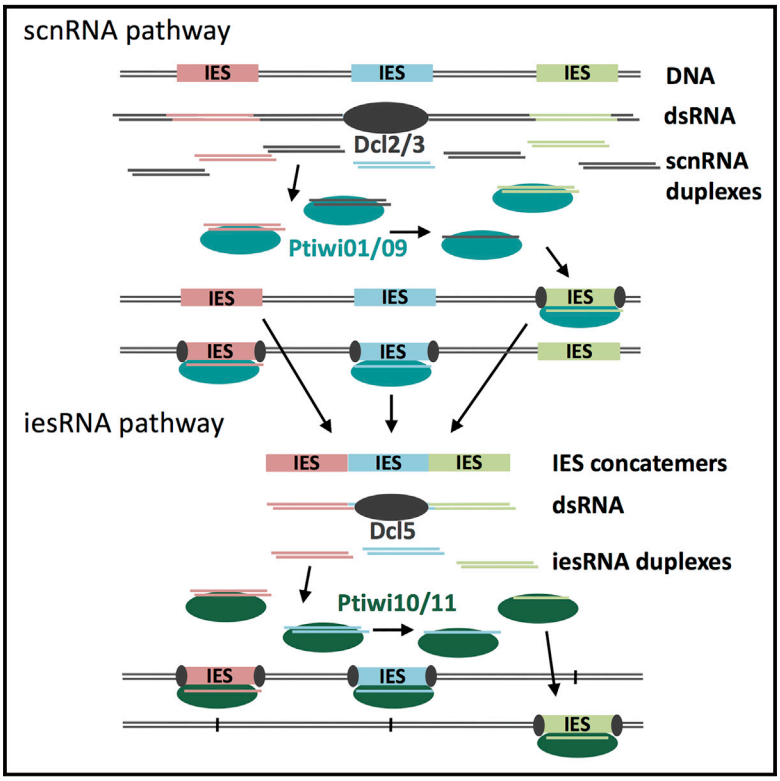
Article

Cell Reports

Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA

Graphical Abstract



Highlights

- Identification of two Piwi proteins (Ptiwi10/11) associated with iesRNAs
- Piwi proteins bind Dicer-produced sRNAs and remove passenger strands
- Ptiwi10 is expressed from the new somatic macronucleus
- DNA elimination activates gene transcription

Authors

Dominique I. Furrer, Estienne C. Swart, Matthias F. Kraft, Pamela Y. Sandoval, Mariusz Nowacki

Correspondence

mariusz.nowacki@izb.unibe.ch

In Brief

Furrer et al. show that two small RNA classes are bound by different Piwi proteins to mediate two functionally separate DNA elimination pathways in *Paramecium*.

Furrer et al., 2017, Cell Reports *20*, 505–520
July 11, 2017
http://dx.doi.org/10.1016/j.celrep.2017.06.050

Cell Reports
Article

Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA

Dominique I. Furrer,[1,2] Estienne C. Swart,[1] Matthias F. Kraft,[1] Pamela Y. Sandoval,[1] and Mariusz Nowacki[1,3,*]
[1]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland
[2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Freiestrasse 1, 3012 Bern, Switzerland
[3]Lead Contact
*Correspondence: mariusz.nowacki@izb.unibe.ch
http://dx.doi.org/10.1016/j.celrep.2017.06.050

SUMMARY

Piwi proteins and piRNAs protect eukaryotic germlines against the spread of transposons. During development in the ciliate *Paramecium*, two Piwi-dependent sRNA classes are involved in the elimination of transposons and transposon-derived DNA: scan RNAs (scnRNAs), associated with Ptiwi01 and Ptiwi09, and iesRNAs, whose binding partners we now identify as Ptiwi10 and Ptiwi11. scnRNAs derive from the maternal genome and initiate DNA elimination during development, whereas iesRNAs continue DNA targeting until the removal process is complete. Here, we show that scnRNAs and iesRNAs are processed by distinct Dicer-like proteins and bind Piwi proteins in a mutually exclusive manner, suggesting separate biogenesis pathways. We also demonstrate that the *PTIWI10* gene is transcribed from the developing nucleus and that its transcription depends on prior DNA excision, suggesting a mechanism of gene expression control triggered by the removal of short DNA segments interrupting the gene.

INTRODUCTION

Piwi proteins, belonging to the Argonaute family, are associated with Piwi-interacting RNAs (piRNAs), a class of small non-coding RNAs (sRNAs) and are involved in the biogenesis, transport, and use of these sRNAs (Luteijn and Ketting, 2013). Argonaute proteins contain four signature domains: the N-terminal, PAZ, MID, and PIWI domains, assembled into two lobes.

The Argonaute protein family can be classified into two subclades, Ago and Piwi. Agos are ubiquitously expressed in most organisms and bind double-stranded small RNAs, such as small interfering RNAs (siRNAs) and microRNAs (miRNAs) that are generated by Dicer cleavage. They are involved in either post-transcriptional gene silencing (PTGS) or transcriptional gene silencing (TGS) (Czech and Hannon, 2011). Piwi proteins, on the other hand, are expressed in animal gonads and bind piRNAs. The precursors of piRNA are single-stranded transcripts from specific regions containing transposable elements (TEs), and processed piRNAs show a strong 5′ U bias (Brennecke et al., 2007).

Together, Piwis and piRNAs mediate silencing of TEs in different eukaryotes, and their function is mainly studied in fruit flies, mice, and humans. Because TEs are mobile genetic elements, they have the ability to change their position in the host genome, including inserting themselves into functional genes, which can have a major effect on their host genome's integrity (Goodier and Kazazian, 2008). Piwi proteins and piRNAs present in germline cells assist in maintaining genome integrity by suppressing transposon movement. Different mechanisms of transcriptional and post-transcriptional silencing that target and prevent TE activation in germline cells have evolved among eukaryotes (Haase, 2016).

Paramecium tetraurelia and other ciliates evolved a unique way to silence mobile elements by eliminating them from their somatic genomes during sexual development. Ciliates have distinct germline and somatic nuclei. *P. tetraurelia* has two diploid, transcriptionally inactive micronuclei (MICs) acting as germline nuclei and one somatic polyploid, transcriptionally active macronucleus (MAC). During every sexual cycle, a new somatic MAC develops from one of the MICs. In the early stages of development, *P. tetraurelia*'s two MICs undergo meiosis; one haploid nucleus produced by meiosis is exchanged with that of its partnered cell (conjugation) and then fused to its sister haploid nucleus (karyogamy). Alternatively, the haploid nuclei directly fuse through a self-fertilization process (autogamy). The newly fused zygotic nucleus divides twice more, and the resultant nuclei then develop into two new MACs and two new MICs. While the new nuclei are developing, the old MAC is fragmented, and, with successive cell divisions, these fragments are rapidly diluted and lost. During the development of the new MAC, the DNA it contains is massively amplified (~800n). Concurrently, the genome is reorganized by genome-wide elimination of repetitive sequences, TEs, and transposon-derived DNA segments (internal eliminated sequences [IESs]).

The *P. tetraurelia* germline genome contains multiple copies of Tc1/mariner transposons and over 45,000 unique IESs (Arnaiz et al., 2012). Most IESs are very short (length mode, 28 bp) and are flanked by 5′-TA-3′ dinucleotides. They are excised from the newly developing somatic genome by the domesticated PiggyBac-related transposase (PiggyMac) (Arnaiz et al., 2012;

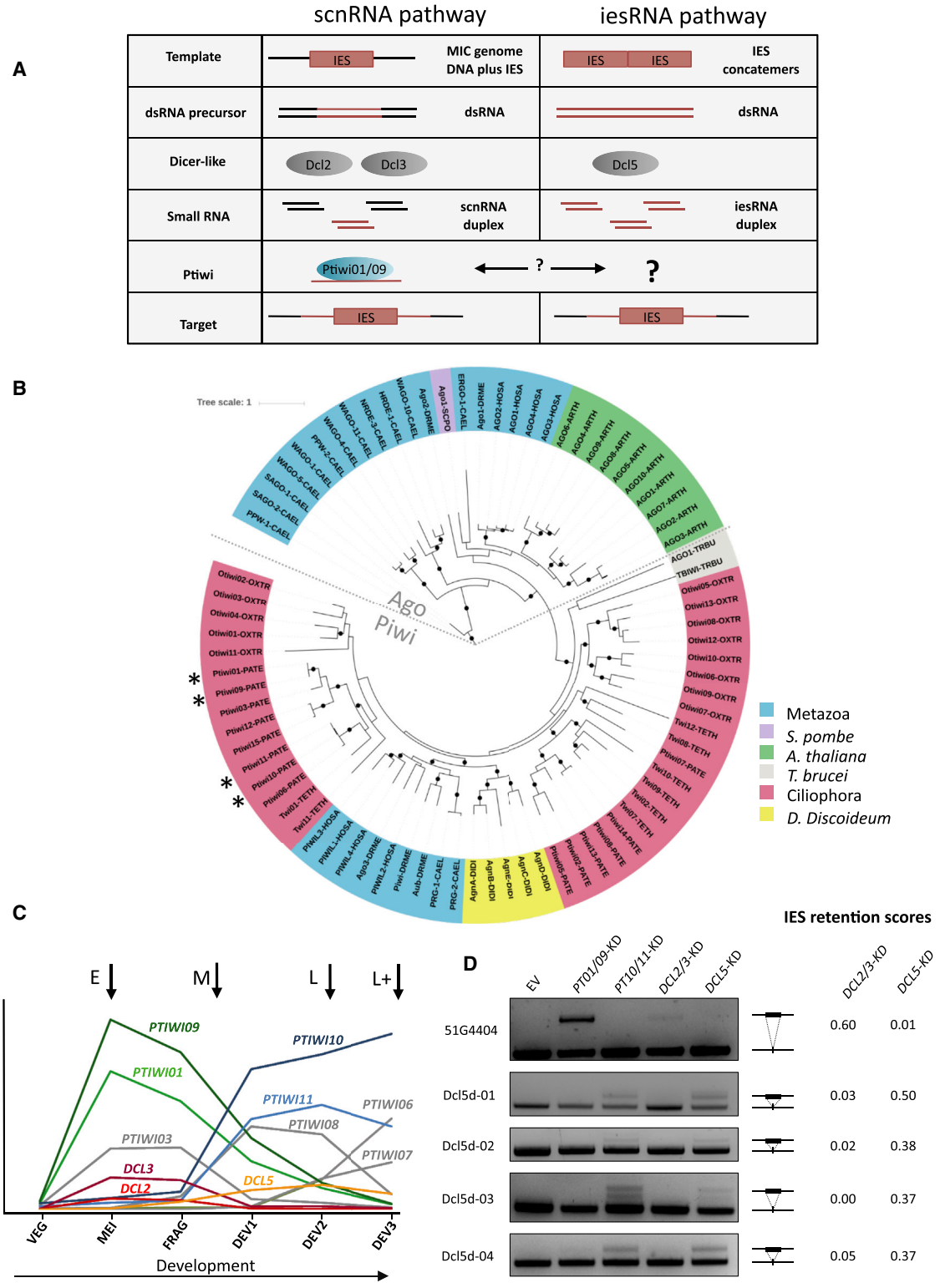


Figure 1. *Paramecium* Piwi Proteins and the Effects of Silencing of Their Genes on IES Excision

(A) Schematic of the scnRNA and iesRNA pathway.

(B) Phylogeny of Ago protein and Piwi proteins from various eukaryotes. ARTH, *Arabidopsis thaliana*; CAEL, *Caenorhabditis elegans*; DIDI, *Dictyostelium discoideum*; DRME, *Drosophila melanogaster*; HOSA, *Homo sapiens*; OXTR, *Oxytricha trifallax*; PATE, *Paramecium tetraurelia*; SCPO, *Schizosaccharomyces*

(legend continued on next page)

Baudry et al., 2009). Repetitive sequences and TEs are excised imprecisely, whereas IES excision occurs in a precise manner, which is necessary because they often interrupt protein coding regions. The end sequences of some IESs are insufficient for their recognition and complete excision by PiggyMac, and it was previously shown that two types of small scan RNAs (scnRNAs and RNAs that exclusively match to IESs [iesRNAs]) are involved in recognizing and triggering the excision of IESs in *P. tetraurelia* (Figure 1A).

scnRNAs are precisely 25 nt long and display a prominent 5′ UNG sequence bias (Lepère et al., 2009; Sandoval et al., 2014). They are generated early during development from long double-stranded RNAs by bidirectional transcription of the entire MIC genome (Lepère et al., 2009). These precursors are cut by two Dicer-like proteins, Dcl2 and Dcl3, forming scnRNA duplexes. One strand of the scnRNAs is then transported into the parental MAC, where the degradation of scnRNAs that are complementary to the parental MAC genome occurs. This process is called ''RNA scanning'' (Mochizuki et al., 2002). The remaining MIC genome-matching scnRNAs are transported into the new MAC, where they target IESs for excision. The second class of development-specific small RNAs appears later during development and consists of iesRNAs, with a strong bias toward their ends. Unlike scnRNAs, iesRNAs are between 23–31 nt long and display a 5′ UAG signature and a weak 3′ CNAUN signature. An additional development-specific Dicer-like protein, Dcl5, is responsible for their production (Sandoval et al., 2014). Because iesRNAs are exclusively IES-matching, it is suggested that they are transcribed after the first round of IES excision. Short excised IESs are ligated randomly to form concatemers, which are then transcribed to long non-coding RNAs (ncRNAs) (Allen et al., 2017). Dcl5 then processes the iesRNA precursors into short iesRNA duplexes. It is proposed that the iesRNAs are targeting the remaining copies of the IESs for excision, like a feedback mechanism, to assure that all IESs are completely removed from the newly developed somatic genome.

In *P. tetraurelia*, 14 different Piwi (Ptiwis) proteins are present, likely predominantly generated by the three successive whole-genome duplications that have occurred in its ancestors (Bouhouche et al., 2011). Even though there are no members of the Ago subclade present in ciliates (Mochizuki et al., 2002; Bouhouche et al., 2011; Fang et al., 2012), RNAi-related pathways are functional. Six of the *Paramecium* Piwis are expressed during vegetative growth function, like typical Agos, and are probably all involved in different RNAi pathways (Bouhouche et al., 2011). Classical post-transcriptional gene silencing was shown to be induced by untranslatable transgenes or by feeding cells with *E. coli* producing double-stranded RNAs (dsRNAs) complementary to the target gene (Ruiz et al., 1998; Galvani and Sperling 2002). Two vegetatively expressed Ptiwi proteins, Ptiwi13 and Ptiwi14, were shown to be involved in these pathways (Bouhouche et al., 2011). Another RNAi pathway has also been demonstrated to exist in *P. tetraurelia* that works together with the highly divergent RNA-dependent RNA polymerase RDR3 to produce sRNAs (Marker et al., 2010; Baranasic et al., 2014). Additionally, recent studies revealed that transgene-induced silencing can induce dynamic heterochromatin remodeling at an endogenous gene locus and that Ptiwi13 and Ptiwi14 are necessary for this (Götz et al., 2016).

The remaining eight Piwi proteins in *Paramecium* are exclusively expressed during development (Bouhouche et al., 2011). Only two of these proteins (Ptiwi01 and Ptiwi09) were studied prior to the present study. It was shown that, in the absence of this pair of proteins, scnRNAs decrease dramatically, leading to retention of certain IESs (Bouhouche et al., 2011).

The major goal of the present study was to identify the *P. tetraurelia* Piwi proteins that interact with iesRNAs and to obtain a better understanding of the pathways and classes of small RNAs involved in genome reorganization. We identified two late-expressed ohnologs, Ptiwi10 and Ptiwi11, as binding partners of iesRNAs. By performing RNA pull-down experiments, we show that scnRNA and iesRNA pathways are distinct. Additionally, our results suggest that the strong sequence bias at the ends of sRNAs and the sRNAs lengths are independent of Piwi binding preferences. In Piwi knockdown experiments, the small RNAs remain double-stranded and keep their strong 5′ UNG signature. This is consistent with the notion that *P. tetraurelia* Dicer-like proteins possess specific sequence cleavage preferences, as suggested by the individual knockdowns of *DCL2* and *DCL3* (Sandoval et al., 2014). Production of Ptiwi10 is enabled by the excision of IESs that interrupt its transcription promoter as well as the coding sequence, which demonstrates a regulatory control mechanism determined at the DNA level. This mechanism allows switching on the iesRNA pathway at a specific developmental stage following the initial step of DNA elimination performed by PiggyMac excisase with the help of scnRNAs.

RESULTS

Paramecium tetraurelia Piwi Proteins and the Effects of Silencing of Their Genes on Dcl5-Dependent IES Excision

All four of the *P. tetraurelia* Piwi proteins investigated in this study, Ptiwi01, Ptiwi09, Ptiwi10, and Ptiwi11, contain the three conserved domains of Argonaute proteins: PAZ, MID, and

pombe; TETH, *Tetrahymena thermophila*; TRBR, *Trypanosoma brucei*. Accession numbers are given in the Supplemental Experimental Procedures. Asterisks indicate the Piwi proteins characterized in this article. Black dotes indicate a bootstrap support of more than 80% of a branching event.

(C) Expression of *PTIWI*s and *DCL*s based on microarray data from Arnaiz et al. (2007, 2010). The cytological stages of the samples collected in this study are indicated by arrows: early (E), in which 30% of the cells have a fragmented old MACs; middle (M), in which all of the cells have fragmented MACs; late (L), in around 80% of the cells new macronuclei were clearly visible; late plus (L+), 6 hr after L.

(D) IES retention was assessed by PCR on DNA obtained after silencing of control (empty vector, EV), *DCL2*/*DCL3* double silencing (*DCL2*/*3*-KD), *DCL5* silencing (*DCL5*-KD), *PTIWI01*/*PTIWI09* double silencing (*PT01*/*09*-KD), and *PTIWI10*/*PTIWI11* double silencing (*PT10*/*11*-KD). IES 51G4404 is a representative epiIES, and Dcl5d-01-04 are IESs whose excision depends on Dcl5. Top bands represent IES^+ PCR products, and bottom band represents PCR products of DNA with the IES excised (IES^-).

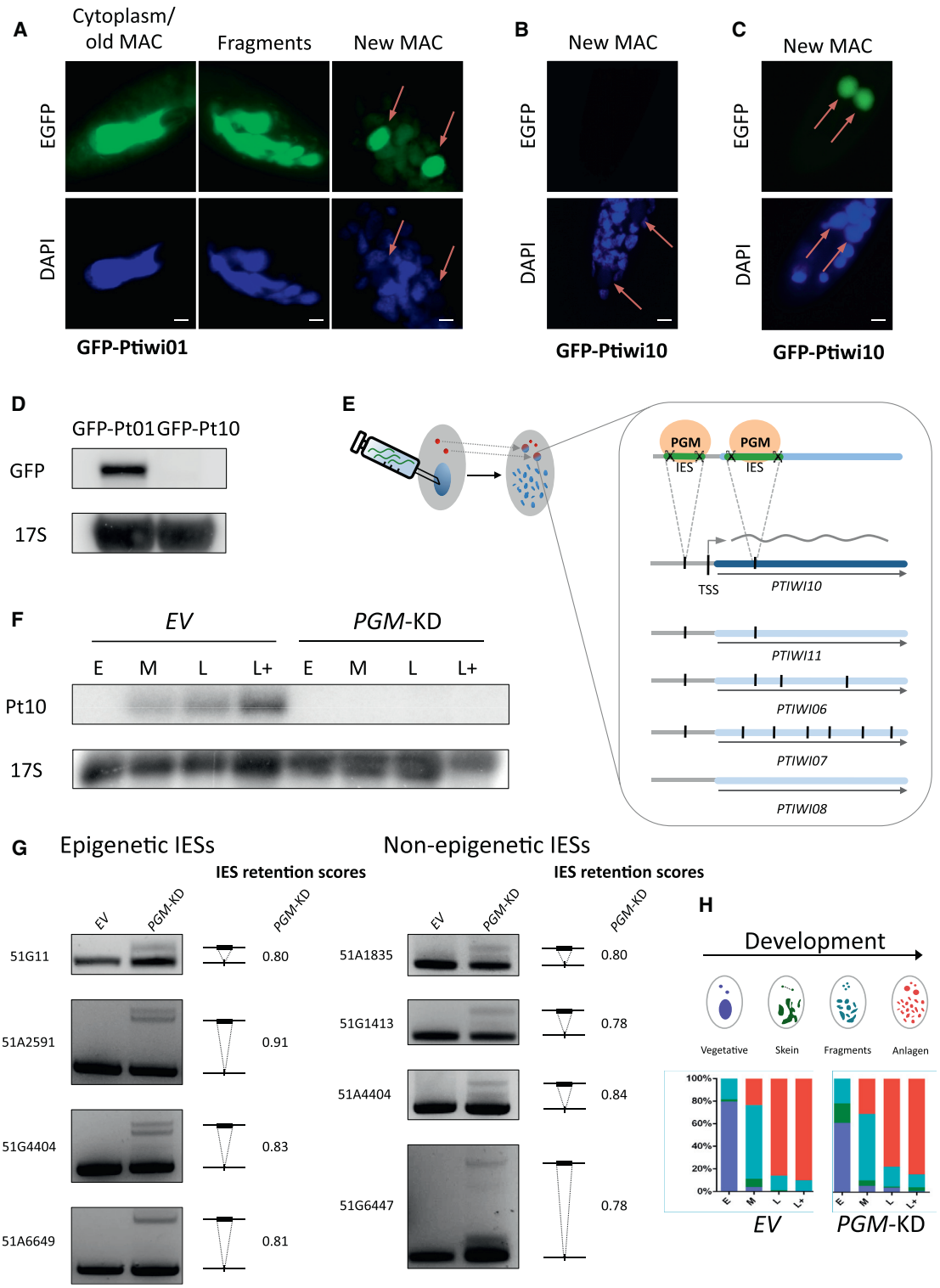


Figure 2. Localization of Ptiwi01 and Ptiwi10 Proteins and Transcription of the *PTIWI10* Gene in New MACs

(A) Subcellular localization of GFP-Ptiwi01 during development. GFP-Ptiwi01 shows the same localization as Ptiwi09-GFP (Bouhouche et al., 2011). At the beginning of development, it localizes to the old MAC and stays there while this nucleus fragments and appears later in the developing new MACs. New MACs are indicated by arrows.

(legend continued on next page)

PIWI. All of the conserved amino acids necessary for small RNA binding are also present in the *P. tetraurelia* Ptiwis as well as the catalytic tetrade responsible for the slicer activity (Yan et al., 2003; Miyoshi et al., 2016; Nakanishi et al., 2012). Ptiwi01 and Ptiwi09 are ohnologs from the most recent whole-genome duplication in *P. tetraurelia* and, at the amino acid level, are 98% identical; Ptiwi10 and Ptiwi11 are also ohnologs and are also 98% identical. In a phylogeny of Argonaute proteins from different eukaryotes, it is clear that Piwi proteins from ciliates, including *P. tetraurelia*, belong to the Piwi subclade (Figure 1B). This is remarkable because all *P. tetraurelia* Ptiwis described to date are involved in RNAi and/or are binding to Dicer/Dicer-like products (Bouhouche et al., 2011; Götz et al., 2016).

Ptiwi03 and the ohnologs Ptiwi01 and Ptiwi09 are expressed early during development. Because iesRNAs peak later during development and are produced in the developing new MACs (Sandoval et al., 2014), we reasoned that Ptiwis with a later peak expression during development might be involved in iesRNA biogenesis and binding and interaction with IESs. Five Ptiwi proteins (Ptiwi06, Ptiwi07, Ptiwi08, and the ohnologs Ptiwi10 and Ptiwi11) are upregulated late during development and were therefore candidates for this (Figure 1C).

To assess which of the late-expressed Ptiwi proteins may be involved in the iesRNA pathway, all of the late-expressed Ptiwis were silenced, and genomic DNA was extracted from cells after development. To assess IES retention following gene silencing, we performed PCRs with primers flanking those IESs, which only show retention in the *DCL5* knockdown (*DCL5-KD*) and not in *DCL2/3* double knockdown (*DCL2/3-KD*) (Dcl5-dependent and Dcl2/3-independent, respectively). IESs were selected based on their previously published retention scores (Sandoval et al., 2014) (IES retention scores for all of the IESs used in the presented study are shown in Table S1). Only double knockdown of *PTIWI10* and *PTIWI11* (*PT10*/11-KD) led to retention of the Dcl5-dependent, Dcl2/3-independent IESs (Figure 1D). Knockdown of the other late-expressed *PTIWI* genes as well as the single knockdowns of *PTIWI10* and *PTIWI11* did not lead to IES retention (data not shown). As a control for this experiment, we analyzed the retention of IES 51G4404, which was previously shown to be dependent on the scnRNA pathway (Sandoval et al., 2014). This IES was retained only after *PT01/09*-KD or *DCL2/3*-KD and not after *PTIWI10/11*-KD or *DCL5*-KD (Figure 1D). The experiment suggests that *PTIWI01/09*-KD affects excision of the Dcl2/3-dependent IESs (epiIESs) but not the Dcl5-dependent IESs (Dcl5d-IESs).

PTIWI10 Is Transcribed in the Developing MAC, and Ptiwi10 Localizes to the Developing MAC

Despite the high similarity (98% at amino acid level) between Ptiwi01 and Ptiwi09, we wanted to examine whether their cellular localization differs. It was previously shown that Ptiwi09-GFP localizes to the parental MAC in early development and that it relocalizes to the developing MACs in late development (Bouhouche et al., 2011). To determine the localization of Ptiwi01, we transformed vegetative *P. tetraurelia* cells with an N-terminal GFP-tagged Ptiwi01 construct that is expressed during development. GFP-Ptiwi01 shows the same localization pattern as GFP-Ptiwi09 (Figure 2A).

Next we examined the localization of N-terminal GFP-tagged Ptiwi10. Despite several attempts, no expression from this construct was detected (Figure 2B). We also did not detect any transcription of the GFP-Ptiwi10 construct in northern blots (Figure 2D). We therefore put the GFP-Ptiwi10 gene construct under the control of a region upstream of *DCL5* expected to contain this gene's promoter because the *DCL5* and *PTIWI10* expression profiles are similar during development (Figure 1C). The new construct allows detection of a strong GFP signal in the two developing MACs (Figure 2C).

We suspect that the reason for the missing transcription of the original GFP-Ptiwi10 construct is that the endogenous gene is exclusively expressed from the developing MACs and that constructs with its endogenous promoter are not recognized by the transcription machinery within the old MAC, where the DNA construct is present after microinjection (Figure 2E). Because it is impossible to microinject DNA into the two new MACs during development (which are very small and visually indistinguishable from the fragments of the old MAC under microinjection conditions), we decided to indirectly test our hypothesis. In the micronuclear version of the genome, we found two IESs whose retention could potentially affect expression of Ptiwi10. One IES is situated within the *PTIWI10* coding sequence (position

(B and C) Subcellular localization of GFP-Ptiwi10 (B) and GFP-Ptiwi10$_{(Dcl5flank)}$ (C) during development. Localization of GFP-Ptiwi10 and GFP-Ptiwi10$_{(Dcl5flank)}$ was monitored during development following transformation of vegetative cells with these constructs. Only GFP-Ptiwi10$_{(Dcl5flank)}$ localized in new MACs (which are indicated by arrows). Scale bars, 10 μm.

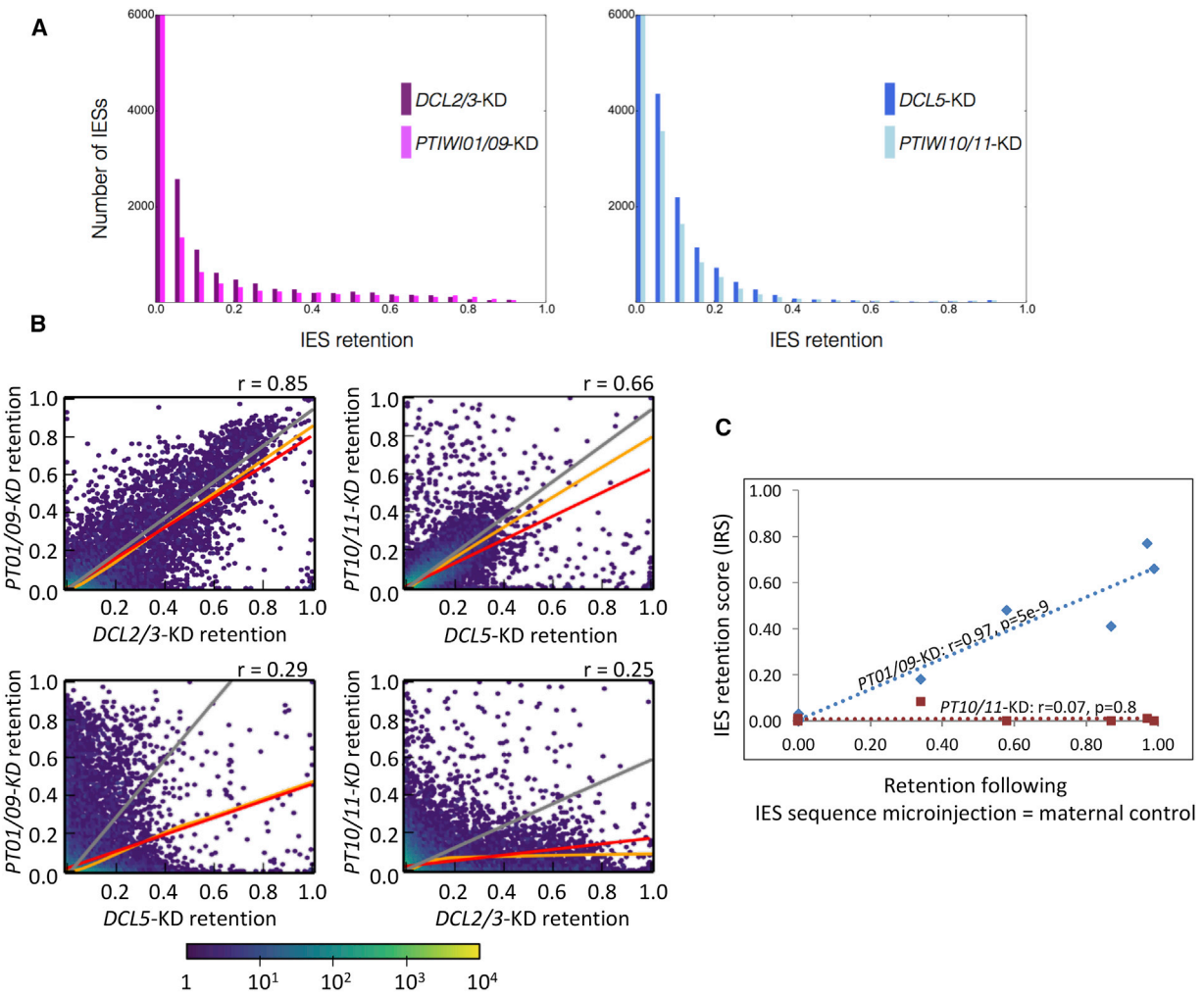
(D) Assessment of Ptiwi gene expression by northern blotting of total RNA obtained from transformed cells. Shown are GFP-Ptiwi01 from an early time point and GFP-Ptiwi10 from a later time point during development. The probe was against the entire GFP sequence, and 17S rRNA was used as a loading control.

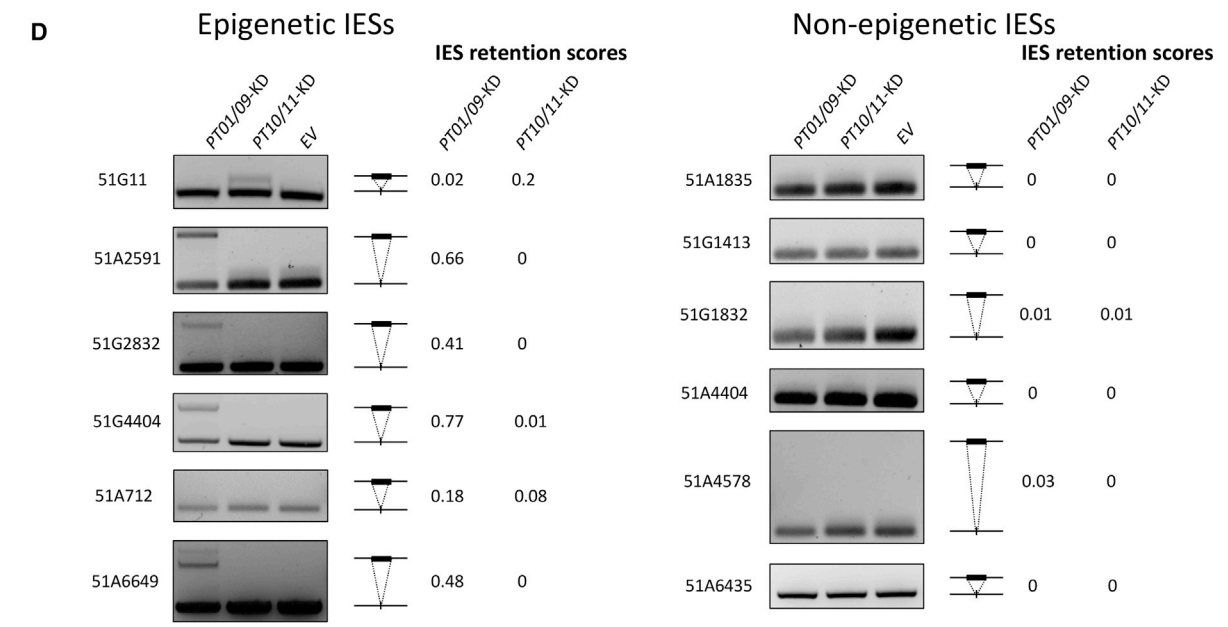
(E) Schematic of the putative effect of *PGM*-KD on *PTIWI10* transcription. When PiggyMac is absent, two IESs, either 48 nt upstream or 442 bp downstream of the start codon, are retained in the genome, which blocks *PTIWI10* transcription. The syringe indicates microinjection into the vegetative MAC. IESs located upstream and inside of the late-expressed *PTIWI* genes are indicated.

(F) Expression of *PTIWI10* as assessed by northern blotting of total RNA obtained from control (EV) cells or cells following *PGM*-KD. Total RNA samples from four different time points were used. A probe against 530 nt of *PTIWI10* and 17S rRNA was used as a loading control.

(G) The efficiency of *PGM*-KD was assessed by performing IES retention PCRs on DNA obtained after silencing of control (EV) and *PGM-KD*. The top band represents IES^{+}, and the bottom band represents DNA with excised IESs (IES^{-}). All tested IESs were partially retained upon PiggyMac silencing.

(H) Distribution of cytological stages of samples collected during cell development from EV and *PGM*-KD cells. A schematic of the nuclear morphology used to characterization the stages is shown above. Bars indicate percentages of cells in a specific stage for each of the collected samples. The cytological stages of the samples collected were as follows: E, in which 30% of the cells have a fragmented old MACs; M, in which all of the cells have fragmented MACs; L, in around 80% of the cells new macronuclei were clearly visible; L+, 6 hr after L. Blue, intact old MAC; green, cells with the old MAC displaying the characteristic skein formation; turquoise, old MAC is completely fragmented; red, visible new MACs.





(legend on next page)

422 nt) and, thus, when retained, would cause a translation frameshift leading to a premature stop codon (position 466 nt) and production of a nonfunctional Piwi protein. In principle, this frameshift may also trigger removal of the mRNA by nonsense-mediated decay. The other IES is located 48 nt upstream of the start codon and could possibly block the transcription of *PTIWI10* by disrupting the promoter region if the IES is not properly excised. Our analysis of the other four late-expressed Ptiwi genes, *PTIWI11*, *PTIWI06*, *PTIWI07*, and *PTIWI08*, revealed that three of them, except *PTIWI08*, have IESs present not only within the coding regions but also in a similar position within the regions expected to contain their transcription promoters (Figure 2E). Thus, these other, late-expressed Ptiwis might also be regulated in a similar way as *PTIWI10*.

To test whether the presence of the two IESs in *PTIWI10* affects the expression of this gene, we silenced the PiggyMac excisase (*PGM*-KD). After *PGM*-KD, *PTIWI10* mRNAs were not detected by northern blotting (Figure 2F; the consistency of the cytological stages of development is shown in Figure 2H, and a description of cell stages can be found in the Experimental Procedures). The effectiveness of *PGM* silencing was indirectly confirmed by IES retention PCRs on genomic DNA extracted from post-autogamous cells after *PGM*-KD (Figure 2G). Thus, we deduce that retention of the IESs because of *PGM*-KD either prevents the transcription of *PTIWI10* in the developing MAC or leads to nonsense-mediated mRNA decay (NMD) of frame-shifted IES-containing mRNAs.

PTIWI01/09-KD Affects Dcl2/3-Dependent IESs, and *PTIWI10/11*-KD Affects Dcl5-Dependent IESs

To analyze the effects of *PTIWI01/09*-KD and *PTIWI10/11*-KD on the entire *P. tetraurelia* IES population, DNA from new MACs was isolated after silencing (knockdown efficiency was assessed by a northern blot, as shown in Figure S1A; for cytological stages of development, see Figure S1B). *PTIWI01/09*-KD affected the excision of only a small proportion of the IESs, but some IESs were strongly retained in this knockdown. In contrast, *PTIWI10/11*-KD affected more IESs overall, but, on average, the effect was slightly weaker in comparison with *PTIWI01/09*-KD (Figure 3A). Compared with the previously published IES retention following silencing of the early-expressed *DCL*s (2/3) (Figure 1C), the IES retention score distribution of the early-expressed *PTIWI*s (01/09) looks very similar, and, likewise, the IES retention score distributions of the late-expressed *PTIWI*s (10/11) and *DCL5* are quite similar (Figure 3A). We observed a good correlation in IES retention scores between *PTIWI01/09*-KD and *DCL02/03*-KD, and so we infer that, like Dcl2/3, Ptiwi01/09 proteins are required for complete excision of epiIESs (Figure 3B). Contrary to a previous report that *PTIWI01/09*-KD also affects non-maternally controlled IESs (IES that were not retained following microinjection of IES sequences identical to them into the old MAC; Duharcourt et al., 1998; Bouhouche et al., 2011), we observed no retention of non-maternally controlled IESs with either IES retention PCRs (Figure 3D) or the higher sensitivity of high-throughput sequencing (Figure 3C). *PTIWI10/11*-KD affects the same IESs as *DCL5*-KD and only affects IES 51G11 within the subset of maternally controlled IESs (Figure 3D). There is only a weak correlation in IES retention scores between *PTIWI01/09*-KD and *DCL5*-KD and between *PTIWI10/11*-KD and *DCL2/3*-KD (Figure 3B). Together, these results suggest that Piwi01/09 interact primarily with the products of Dcl2/3 and that Ptiwi10/11 interact primarily with the products of Dcl5.

PTIWI01/09-KD and *PTIWI10/11*-KD Lead to scnRNA and iesRNA Depletion, Respectively

To study the effect of the abovementioned *PTIWI01*/09-KD and *PTIWI10/11*-KD on the small RNA population, small RNAs from different developmental stages were analyzed by gel electrophoresis and deep sequencing. In control cells treated with an empty vector (EV), the early developmental time point shows a prominent 25-nt scnRNA peak mostly matching MAC genome-destined DNA (produced from MIC genome transcripts early during development) (Figure 4A, *EV*-E); later, IES-matching iesRNAs appear, as described previously (Sandoval et al., 2014; Figure 4, *EV*-L and *EV*-L+). After depleting the early-expressed Ptiwi01 and Ptiwi09, the 25-nt-long scnRNA population at the early time point (*PTIWI01/09*-KD-E) decreased 4.5-fold compared with the scnRNAs from the early time of the control (*EV*-E). iesRNAs levels were also reduced at the two late time points following *PTIWI01/09*-KD (*PTIWI01/09*-KD-L and *PTIWI01/09*-KD-L+). This effect is consistent with a model in which iesRNAs are produced from excised IESs (Sandoval et al., 2014). *PTIWI10/11*-KD reduces the iesRNA population (26–29 nt); e.g., 8.2-fold at the L+ time point compared with the control (*PTIWI10/11*-KD-L+). The absolute decreases of the small RNAs visualized by loading the total RNA samples on denaturing polyacrylamide gel reflect the

Figure 3. Effects of *PTIWI01/09*-KD and *PTIWI10/11*-KD on IES and Transposon Excision

(A) Genome-wide IES retention measured from Illumina sequencing of DNA from newly developed MACs for *DCL2/3* double-silenced cells (*DCL2/3*-KD), *DCL5* silenced cells (*DCL5*-KD) (Sandoval et al., 2014), *PTIWI01* and *PTIWI09* double-silenced cells (*PTIWI01/09*-KD), and *PTIWI10* and *PTIWI11* double-silenced cells (*PTIWI10/11*-KD). For IESs with 0%–5% IES retention, the bars are truncated. The maximal values for these bars are as follows: *DCL2/3*-KD, 37,421; *DCL5*-KD, 35,033; *PTIWI01/09*-KD, 39,982; *PTIWI10/11*-KD, 37,198.

(B) Relationships in IES retention among knockdown pairs of *PTIWI01/09*-KD, *PTIWI10/11*-KD, *DCL2/3*-KD, and *DCL5*-KD. Hexagonal binning of IES retention scores was used to generate the plots. Pearson's correlation coefficients are given above each subgraph. Red lines are for ordinary least-squares (OLS) regression, orange lines for LOWESS, and gray lines for orthogonal distance regression (ODR).

(C) To examine the relationship between maternal control and IES retention scores (IRSs), maternal control was calculated as the maximal observed retention described previously (Duharcourt et al., 1998; Sandoval et al., 2014) with the IRS calculated in the present study (both maternal control scores and IRS are given in Table S1). Dashed lines corresponding to linear regressions are shown along with Pearson's r and the two-tailed p value for hypothesis testing with a null hypothesis of a regression slope of zero.

(D) IES retention PCRs were performed on the same IESs as those used in (C). The same new MAC DNA was used as that in the Illumina sequencing of new MAC DNA and transposon probing. The top band represents the IES^{+} form; the bottom band represents DNA with excised IESs (IES^{-}).

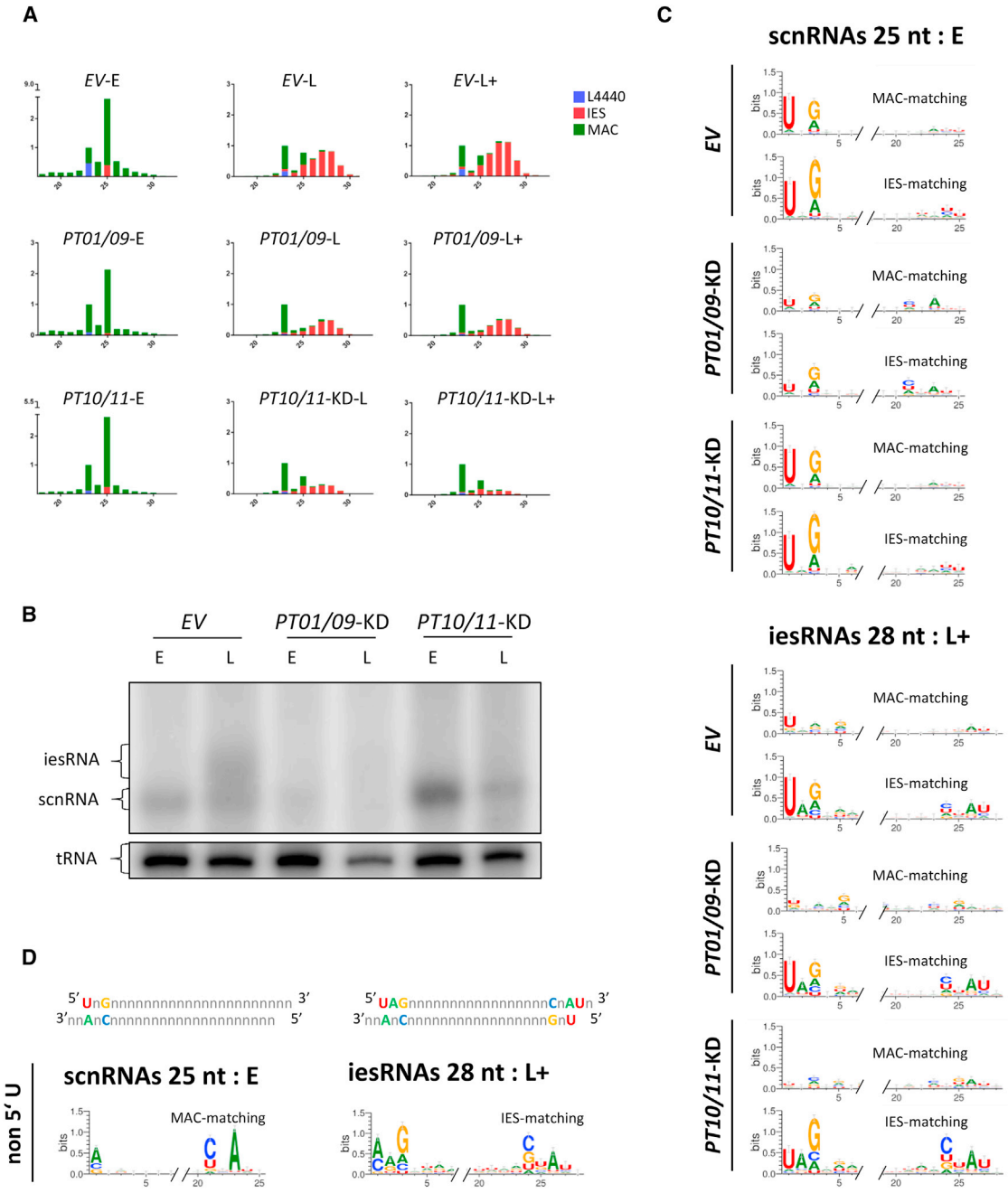


Figure 4. Effect of *PTIWI01/09*-KD and *PTIWI10/11*-KD on Development-Specific Small RNAs

(A) Histogram of sRNA of total 19- to 31-nt small RNA library reads mapped to the macronuclear genome (MAC, green), internal eliminated sequences (IESs, red), and the plasmid backbone used for silencing (L4440, purple). E, L, and L+ development correspond to successive developmental times (for the distribution of cytological stages, see Figure S1B). Reads were normalized to total 23-nt reads, which are predominantly comprised of siRNAs.

(B) Radioactively 5′ end-labeled total RNA from control (EV), *PTIWI01/09*-KD, and *PTIWI10/11*-KD cells was loaded on a 16% denaturing polyacrylamide gel. scnRNAs and iesRNAs are indicated, and tRNA was used as a loading control.

(C) Representative sequence logos of 25-nt and 28-nt sRNAs from control (EV), *PTIWI01/09*-KD, and *PTIWI10/11*-KD cells.

(D) Representative sequence logos of non-5′ U 25-nt scnRNAs from *PTIWI01/09*-KD cells and non-5′ 28-nt iesRNAs from *PTIWI10/11*-KD.

relative decreases in sRNAs determined by high-throughput sequencing (Figure 4B).

By analyzing the 5′ signature of the 25-nt sRNAs obtained after knockdown, we observed that the 5′ UNG signature of the 25-nt sRNAs is weaker (Figure 4C) in *PTIWI01/09*-KD than in control cells, and a weak CNANN signature appears at the 3′ end. Because normally Argonaute proteins select the guide strand and remove the passenger strand, it is possible that the

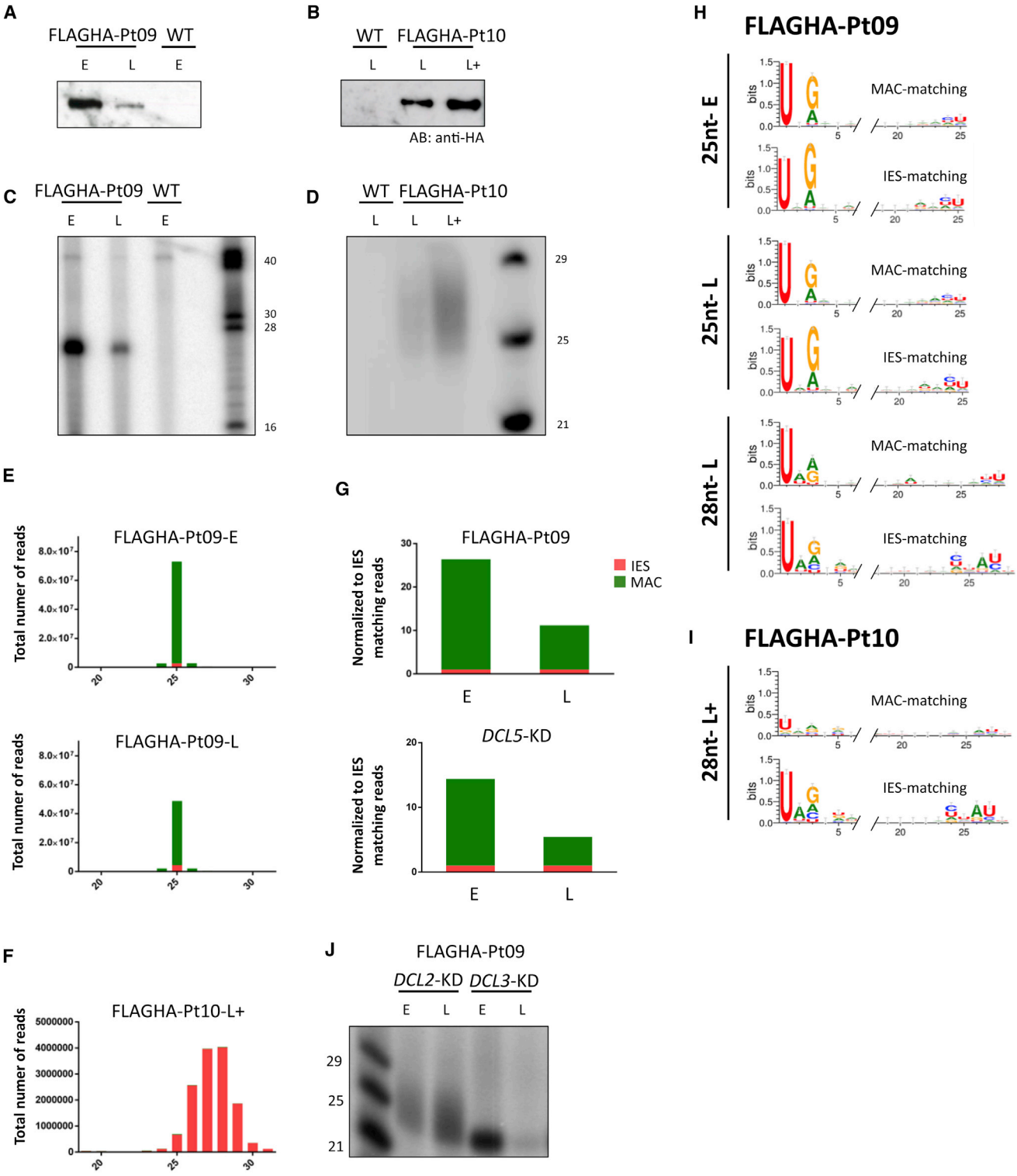


Figure 5. Small RNAs from Different Time Points Obtained from RIP of FLAGHA-Ptiwi09 and FLAGHA-Ptiwi10 Proteins Expressed in *Paramecium* Cells

(A and B) Western blot of FLAGHA-tagged Ptiwi09 (A) and Ptiwi10 (B) pulled down with anti-HA antibodies.

(C and D) Radioactively 5′ end-labeled small RNAs associated with FLAGHA-tagged Ptiwi09 (C) or Ptiwi10 (D).

(E) Stacked histogram of FLAGHA-Ptiwi09 RIP sRNA reads mapped to the MAC (green) and IESs (red) from two different time points.

(legend continued on next page)

scnRNA population remains double-stranded in the absence of Ptiwi01/09. This feature of the *Paramecium* Piwi proteins is quite unique because Piwis in animals bind single-stranded RNA (ssRNA) precursors and, therefore, do not need to eliminate passenger strands. Next we analyzed the scnRNAs that do not show the 5′ U of the guide strand. Indeed, our analysis of the non-5′ U scnRNAs reveals a prominent CNANN at their 3′ end, reflecting the signature of the passenger strand (Figure 4D). In the case of iesRNAs, the result is not so clear because of their palindromic nature; i.e., both the guide and the passenger strands of iesRNAs have both 5′ UAG and 3′ CNAUNN signatures (Figure 4D).

Ptiwi09 Binds scnRNAs, and Ptiwi10 Binds iesRNAs

To confirm that Ptiwi09 binds scnRNAs and Ptiwi10 binds iesRNAs, RNA immunoprecipitation (RIP) was performed (Figure 5; Figure S2A). Ptiwi09 RIP samples from both the early and late developmental time points revealed the presence of sRNAs that were 25 nt long (the expected size of scnRNAs; Figure 5C), and the sRNAs bound to Ptiwi10 ranged from 24–31 nt (the expected size of iesRNAs; Figure 5D).

High-throughput sequencing supports the specificity of Ptiwi09 for 25-nt scnRNAs (Figure 5E). At both time points, scnRNAs showed the characteristic 5′ UNG signature (Figure 5H). At the late time point, where iesRNAs are present in the cell, there are just traces of longer small RNAs bound to Ptiwi09 (only 2.2% of the reads were between 27–31 nt). Additionally, only 13.7% of the 27- to 31-nt reads were IES-matching (i.e., at most, ~0.03% of Ptiwi09-bound small RNAs are iesRNAs). This tiny fraction of IES-matching, 28-nt small RNAs shows the characteristic 5′ UAG and the 3′ CNAUN signatures of iesRNAs (Figure 5H). Small RNAs bound to Ptiwi10 almost exclusively map to IESs (98.5%) and show the characteristic 5′ UAG and the 3′ CNAUN signature of iesRNAs (Figures 5F and 5I). The size of the Ptiwi10 RIP small RNAs, ranging between 24 nt and 31 nt, peaking at 27 nt and 28 nt, reflect the profile of iesRNAs from total small RNA sequencing (Figure 4A).

Because the current RNA-scanning model proposes that the MAC genome-matching scnRNAs are degraded in the old MAC and only the IES-matching scnRNAs are transported to the new MAC, we compared the ratio of IES:MAC genome-matching scnRNAs from the two Ptiwi09 RIPs (Figure 5G). After normalizing all of the reads obtained from the Ptiwi09 RIP to their IES-matching reads, we observe a modest 2.5-fold decrease of the MAC genome-matching scnRNAs at the later time point, but a large proportion of MAC genome-matching scnRNAs still remains present. One possible explanation for why this decrease is so modest is that most of the MAC genome-matching scnRNAs are still present in the old MAC, or, alternatively, in the case of the Ptiwi09 RIP, overexpression of this protein might prevent degradation of these scnRNAs. However, a similar modest decrease (3-fold) in MAC genome-matching scnRNAs can be observed by analyzing reads obtained from *DCL5*-KD cells (no 25-nt-long iesRNAs present), as published by Sandoval et al. (2014) (Figure 5G).

We next investigated whether Ptiwi09 has a preference for binding scnRNAs of a specific size (25 nt). To do so, we performed a Ptiwi09 RIP experiment with cells in which *DCL2* was silenced. *DCL2*-KD leads to a pronounced reduction in 25-nt scnRNAs, leaving behind scnRNAs with a broad range of sizes (from 24–31 nt) (Sandoval et al., 2014). We find that Ptiwi09 does not have a preference for binding only 25-nt sRNAs but, rather, binds all of the sizes generated by Dcl3 (Figure 5J). Conversely, in the case of *DCL3*-KD, which does not change the size of scnRNAs (25 nt), the small RNAs associated with Ptiwi09 are exactly 25 nt long (Figure 5J).

Transposons and the Small RNAs Involved in Their Excision

Because Piwi proteins are involved in suppressing transposon gene expression in eukaryotes, including ciliates (via DNA elimination), we examined whether the development-specific Ptiwis and the corresponding small RNAs are involved in the removal of transposons. Unlike IESs, transposons are imprecisely excised during development, but the exact mechanism and the difference between imprecise and precise elimination are not yet known. To investigate the possible role of Piwi protein in transposon elimination, we performed a Southern blot analysis on the same DNA as used for the analysis shown in Figure 3.

Only *PTIWI01*/09-KD, but not *PTIWI10*/*11*-KD, affects the excision of the two known *P. tetraurelia* transposon classes *Sardine* and *Thon* (Figure 6A), consistent with dependence of transposon elimination on scnRNAs but not on iesRNAs, as shown previously (Sandoval et al., 2014).

Like transposons, many IESs do not seem to require iesRNAs for their excision (Figure 3). However, iesRNAs are still generated from these iesRNA-independent IESs (Sandoval et al., 2014). To investigate whether iesRNAs are produced from transposons during development, we mapped scnRNAs (5,781,331 reads) and iesRNAs (5,709,855 reads) obtained from the Ptiwi09 or Ptiwi10-RIP experiments, respectively, to several different copies of imprecisely excised *Sardine* transposons. As a comparison, we mapped the same amount of small RNAs to the precisely excised transposons remnants called *Anchois* (Arnaiz et al., 2012) and to a concatemer of the eight longest IESs, which have an IES retention score (IRS) of less than 0.01 in *PTIWI10*/*11*-KD (representing iesRNAs-independent IESs).

Only 159 iesRNAs (28 nt) map to the eight annotated *Sardine* transposons (Figure 6B), compared with 2,998 iesRNAs (28 nt) that map to the eight longest IESs (Figure 6C) and the 4,009 iesRNAs (28 nt) that map to the *Anchois* transposons (Figure 6D). On the other hand, scnRNAs are produced of all the elements because 3,507 scnRNAs (25 nt) map to the *Sardine* transposons, 2,850 to *Anchois*, and 2,585 to the eight longest IESs (Figures 6B–6D).

(F) Stacked histogram of FLAGHA-Ptiwi10 RIP sRNA reads mapped to the MAC (green) and IESs (red) from the latest time point collected during development.
(G) Stacked bar plot of FLAGHA-Ptiwi09 RIP sRNA reads and reads of *DCL5*-KD (reads published in Sandoval et al., 2014). On both blots, reads are normalized to the IES-matching reads.
(H and I) Representative sequence logos of small RNAs obtained from FLAGHA-tagged Ptiwi01 (H) or Ptiwi10 RIP (I).
(J) Radioactively 5′ end-labeled small RNAs associated with FLAGHA-tagged Ptiwi09 obtained from *DCL2*-KD and *DCL3*-KD cells.

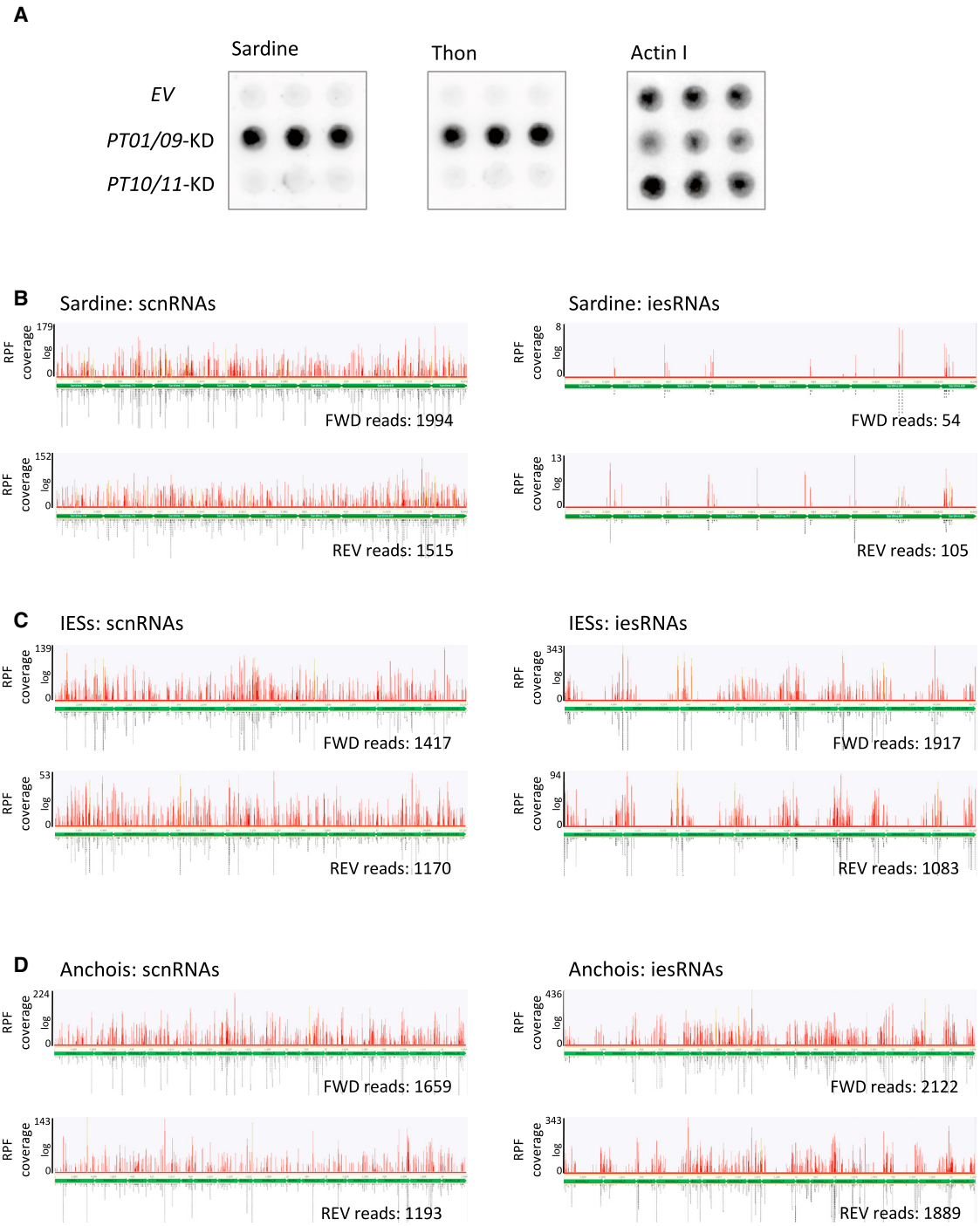


Figure 6. Elimination of Transposons and the Involved Small RNAs

(A) The effect of *PTIWI01/09*-KD and *PTIWI10/11*-KD on transposon excision. Retention of *Sardine* and *Thon* transposons was analyzed using probes specific to them, and a probe against a *P. tetraurelia* Actin 1 gene was used as a loading control. DNA was loaded in triplets.

(B–D) Small RNAs associated with FLAGHA-Ptiwi09 and FLAGHA-Ptiwi10 mapped to *Sardine* transposons, *Anchois* IESs, or the eight longest IESs (concatenated in each case for illustration convenience; IRS of *PTIWI10/11*-KD > 0.01). 3,509 scnRNAs (25 nt) (forward [FWD]-reads, 1,994; reverse [REV]-reads, 1,515) obtained from Ptiwi09 RIP mapped to *Sardines*, 2,852 (FWD-reads, 1,659 + 1,193) to *Anchois*, and 2,581 (FW-reads,: 1,417; REV-reads, 1,170) to the IESs. 159 iesRNAs (28 nt) (FWD, 54; REV-reads, 105) obtained from Ptiwi10 RIP mapped to *Sardines*, 4,011 (FWD-reads, 2,122; REV-reads, 1,889) to *Anchois*, and 3,000 (FWD-reads, 1,917; REV-reads, 1,083) to the IESs. The numbers of reads used during mapping were as follows: scnRNAs, 5,781,331; iesRNAs, 5,709,855. The lengths of the concatemers were as follows: *Sardine* transposons, 49,570 nt; *Anchois* transposons, 42,440 nt; and the eight longest IESs, 33,107 nt.

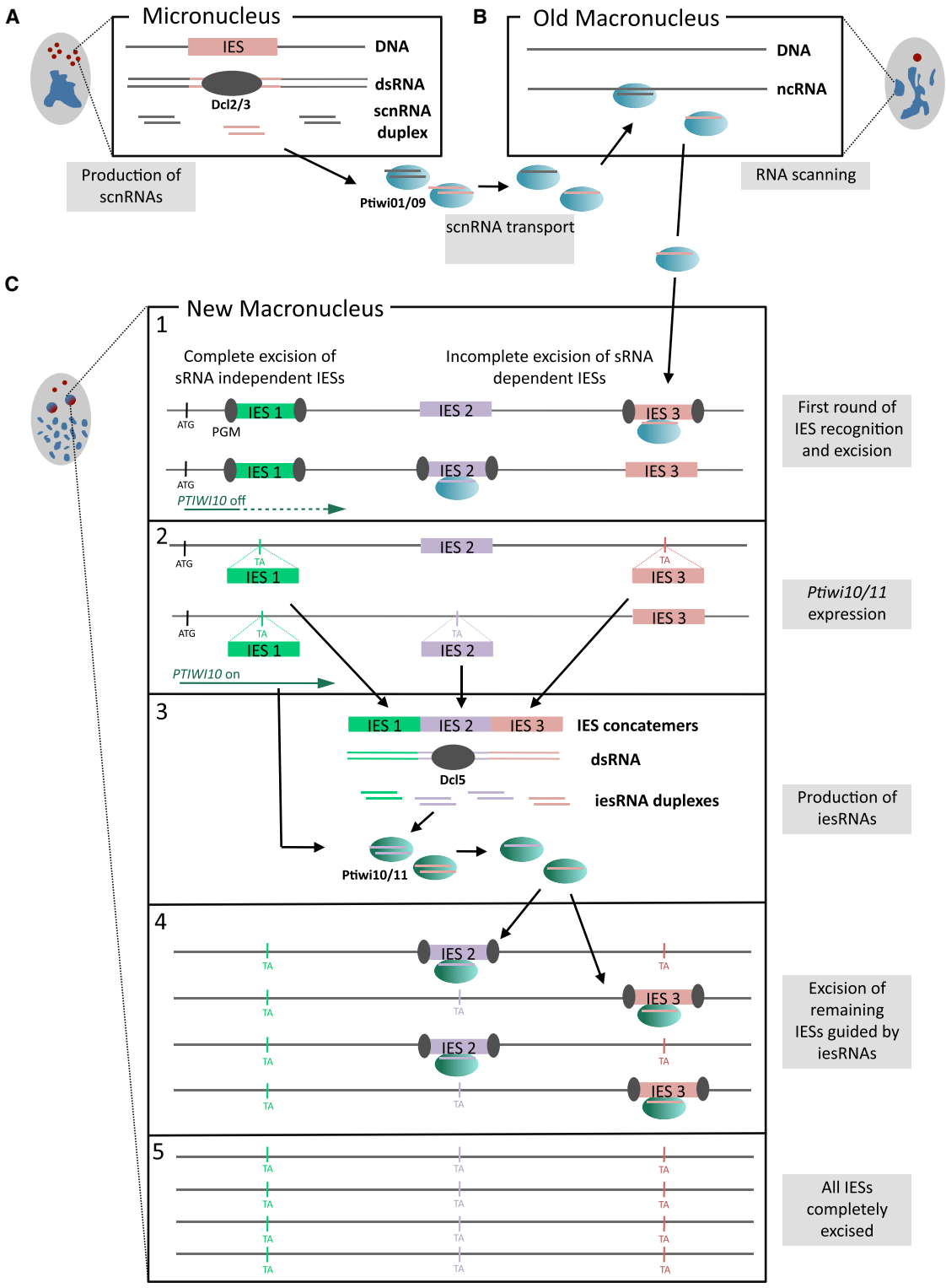


Figure 7. Proposed Model of the Roles of Ptiwi01, Ptiwi09, Ptiwi10, and Ptiwi11 in IES Elimination during Development, Building on Existing Models

(A) During early development, *Paramecium tetraurelia's* germline genome is bi-directionally transcribed. The resulting dsRNAs are cleaved by Dcl2/3 into 25-nt precursor scnRNAs. The precursor scnRNAs are then loaded onto Ptiwi01 and Ptiwi09 proteins, which select the guide RNA strands and remove the passenger strands.

(legend continued on next page)

A possible explanation for why there are so few *Sardine*-derived iesRNAs could be that a different mechanism is being employed for the imprecise excision of transposons that does not lead to post-excision iesRNA production. Because iesRNAs tend to accumulate at IES boundaries (Sandoval et al., 2014), one alternative explanation is that these elements may be embedded within eliminated DNA that has very distantly separated excision boundaries.

DISCUSSION

During *Paramecium tetraurelia* somatic genome development, two distinct classes of small RNAs, scnRNAs and iesRNAs, are involved in excision of transposons and IESs, which are transposon-derived. The dsRNA precursors of scnRNAs are produced in meiotic germline nuclei through the action of Dcl2 and Dcl3, whereas iesRNAs appear later, and their dsRNA precursors are processed by Dcl5. In the present study, we wished to know whether the two small RNA biogenesis pathways may be linked by Piwi proteins that complete their processing and then allow them to target DNA for elimination. We identified two new Piwi proteins necessary for complete IES excision, the ohnologs Ptiwi10 and Ptiwi11. These proteins select iesRNA guide strands and degrade their passenger strands prior to the targeting of DNA excision. Additionally, we provide evidence that the two early-expressed proteins, Ptiwi01 and Ptiwi09, are responsible for scnRNA accumulation. As with a previous report from the knockdown of *DCL2/3*, which shows that scnRNAs are involved in excision of epiIESs (Sandoval et al., 2014), knockdown of *PTIWI01/09* also affects only the excision of epiIES. This is consistent with Dcl2/3 and Ptiwi01/09 working in the same pathway.

Based on our RIP experiments, we can conclude that Ptiwi09 binds only scnRNAs and Ptiwi10 only iesRNAs, even though there is an overlap between the two sRNA classes in their length and sequence (many iesRNAs look like scnRNAs and vice versa). Our result from the *DCL2*-KD experiment shows that Ptiwi09 does not have a length preference for the bound sRNAs (Figure 5J), which makes it similar to Ptiwi10 in this regard.

The RIP experiments were performed with N-terminal FLAG-hemagglutinin (HA) fusion proteins, which is a commonly used method. We cannot exclude the possibility that the presence of the additional 32 amino acids affects the function of the proteins and, therefore, their sRNA binding efficiency. This, however, is unlikely to affect our conclusions because the presented sRNA analysis is qualitative rather than quantitative.

The seemingly high specificity between the Piwi proteins and their associated class of sRNAs may be attributed to the fact that, at the time when scnRNAs are produced, the only available Piwi proteins are Ptiwi01/09. Later during development, when iesRNAs are produced, even though Ptiwi01/09 are present in the same place, the only available Ptiwis are Ptiwi10/11. This would also suggest that there is virtually no exchange of sRNAs between the two types of Piwi proteins. The role of the four Piwi proteins in programmed DNA elimination is summarized in Figure 7.

Our results also suggest that the specific 5′ end sequence signature may be due to Dicer-like protein cleavage preference because, in the absence of Ptiwis, 95% of the 25-nt scnRNAs have a U at the 5′ end of the guide strands or ANN at the 3′ ends of the passenger strands (Figures 4C and 4D).

PTIWI10 is the first gene described in *P. tetraurelia* which is transcribed from the new MAC during development of the new MAC. Until now it has been thought that transcription in the new MAC starts after development is completed. The reason for this idea was that, when a ''non-functional'' new MAC genome is formed, because of the depletion of factors involved in genome rearrangements, the cell does not die immediately (for example, PiggyMac; Baudry et al., 2009) but is able to survive during two to three subsequent vegetative divisions. This is likely because the fragments of the parental MAC are still present in the cell for some time after sexual reproduction, and gene transcription from those fragments is sufficient for cell survival. As the fragments are diluted out after several divisions, the newly formed MAC takes over the gene expression function. The expression of Ptiwi10 (and, presumably, also Ptiwi11) from the developing MAC suggests a mode of gene expression control of genes involved in genome development, which is intimately related to the development dynamics. The late sRNA pathway (iesRNA), which is essential for complete DNA elimination, is only switched on when IESs interrupting the promoter and the coding sequence of Ptiwi10/11 are eliminated by the excision machinery (Figure 7). In other words, the iesRNA pathway will remain inactive unless the early steps of DNA elimination are functional. This may represent a more general mechanism in which genes are unblocked at the right time in the right place through DNA elimination in the developing macronucleus.

The following experiments support the existence of such a mechanism. We show that *PTIWI10* cannot be expressed from the IES-free parental MAC, which we deduced from the lack of expression of transgenic *GFP-PTIWI10* under the original promoter (Figures 2B and 2D). That means that the Ptiwi10 promoter is inactive in the parental MAC. We have also shown that IES excision by PiggyMac excisase is required for the transcription of Ptiwi10 (Figure 2F).

The presence of either of the two IESs in *PTIWI10*, one in the promoter and one in the coding region, would likely lead to the

(B) The scnRNA-Ptiwi complexes localize to the parental MAC, where RNA scanning takes place, and MAC genome-matching scnRNAs are filtered out after binding to putative long ncRNAs. Ptiwi01- and Ptiwi09-bound MIC genome-specific scnRNAs then migrate to the newly developing MAC, were they target IES excision.

(C)1: short non-epiIES, represented by IES1, are very efficiently excised by PiggyMAC (PGM), even in the absence of small RNAs. scnRNA-dependent IESs are, at the same time, incompletely excised. 2: the initial excision of IESs unblocks *PTIWI10/11* genes and, thus, activates the iesRNA pathway. 3: after the first round of IES excision by PiggyMAC, the excised IESs are ligated together to form concatemers and/or IES circles (only concatemers are shown), which enables their bidirectional transcription. Dcl5 then processes the long dsRNA precursors into iesRNA duplexes. When Ptiwi10/11 are expressed, the iesRNA precursors are loaded onto them. This is followed by guide strand selection and passenger strand removal. 4: Ptiwi10/11 associated with iesRNAs target the remaining copies of IESs for excision. 5: new MAC DNA after complete IES elimination. TA dinucleotide (TA) represents the positions from where the IESs were excised.

disappearance of *PTIWI10* mRNA. The IES in the promoter would disrupt transcription initiation, and the one in the coding region would cause a frameshift and a premature stop codon, which would lead to NMD-mediated mRNA degradation. In this case, IES excision would be essential for *PTIWI10* mRNA expression.

The regulation of *PTIWI10* expression may represent a developmental checkpoint in *Paramecium*. The initial step of DNA elimination is performed by PiggyMac excisase through targeting IESs that do not require small RNAs for their recognition and IESs that are scnRNA-dependent (Figure 7C, 1). This initial step removes IESs that disrupt the *PTIWI10* gene and allows the expression of Ptiwi10 (Figure 7C, 2) and the production of iesRNAs from the initially excised IESs (Figure 7C, 3). Next, iesRNAs bound to Ptiwi10/11 target the remaining unexcised IESs (Figure 7C, 4). In this work, we show that the iesRNA pathway can only be initiated when the first round of excision is completed, thus allowing the cell to enter the next developmental stage at the right moment.

EXPERIMENTAL PROCEDURES

Paramecium tetraurelia Cultivation

For all experiments, *Paramecium tetraurelia* strain 51, mating type 7 was used. Cells were grown in wheat grass powder (WGP) medium (Pines International, Lawrence, KS) bacterized with *Klebsiella pneumoniae* and supplemented with 0.8 mg/L of β-sitosterol (Merck). Cultivation was carried out at 27°C as described previously (Beisson et al., 2010).

Cytological Stages of Development

The developmental stages used in our experimental assessment are based on the characterization of cytological stages. In this study, four time points were used: early (E), in which 30% of the cells have a fragmented old MAC; middle (M), in which all of the cells have fragmented MACs; late (L), in around 80% of the cells new macronuclei were clearly visible; and late plus (L+), 6 hr after L.

Gene Silencing

Gene silencing during autogamy was performed by bacterial feeding as described previously (Nowacki et al., 2005). Different regions of the target coding sequences were cloned into the L4440 plasmid (*PTIWI01*, 1,394 bp; *PTIWI09*, 1,396 bp; *PTIWI10*, 1,489 bp; *PTIWI11*, 979 bp). To silence *DCL2*, *DCL3*, and *DCL5*, previously designed silencing constructs were used (Lepère et al., 2009; Sandoval et al., 2014). For *PGM* silencing, the published construct was used (Baudry et al., 2009). The negative control for silencing was performed using the empty L4440 plasmid. The plasmids were transformed into the HT115 (DE3) *Escherichia coli* strain. The primers used are listed in the Supplemental Experimental Procedures.

DNA Extraction

Total genomic DNA extraction for IES PCRs was performed with the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich). For deep sequencing DNA from the new MAC was isolated as previously described (Arnaiz et al., 2012).

GFP Tagging, Injection, and Imaging

Ptiwi01 and Ptiwi10 were N-terminally tagged with a GFP codon optimized for the *P. tetraurelia* genetic code (Nowacki et al., 2005). The GFP-Ptiwi01 construct used the 5′ and 3′ sequences flanking *P. tetraurelia*'s endogenous PTIWI01 (270 bp upstream and 162 bp downstream). GFP-Ptiwi10 used the sequences flanking *P. tetraurelia*'s endogenous *PTIWI10* (260 bp upstream and 36 bp downstream) or *DCL5* (352 bp upstream and 108 bp downstream, as published by Sandoval et al., 2014).

The constructs were linearized and microinjected into the vegetative MAC. Positively injected clones were selected based on confirmation of injected DNA by dot blot analysis. Microscopic images for GFP-Ptiwi01 and GFP-Ptiwi10 localization were taken with a Leica AF6000 system, and GFP-Ptiwi10$_{(Dcl5flank)}$ images were taken with the Axio Vert.A1.

RNA Extraction

RNA from 400 mL of harvested cells at a concentration of 3,000 cells/mL was resuspended in 6 mL of TRI reagent (Sigma-Aldrich). Total RNA extraction was carried out following the TRI reagent BD protocol.

Small RNA Sequencing

Small RNA libraries were produced and sequenced according to standard Illumina protocols, and sequencing was performed on an Illumina HiSeq2500 machine by Fasteris SA.

sRNA Sequencing Mapping

sRNA reads were mapped as described in Sandoval et al., 2014. For knockdown experiments, reads were normalized to the total of 23-nt reads mapping to IESs, the MAC genome, and the L4440 plasmid.

Sequence Logo Generation

Sequence logos were created with WebLogo, version 3.2. The same background base frequencies G = C: 0.30 were used to produce total and MAC genome-matching logos, and G = C: 0.20 were used for IES-matching sRNAs.

RIP and Western Blot

The RIP protocol from Cora et al. (2014) was adapted for use in *P. tetraurelia*.

Ptiwi09 and Ptiwi10 were N-terminally tagged with 3× FLAG-HA tags (FLAGHA) codon-optimized for the *P. tetraurelia* genetic code. The FLAGHA-Ptiwi09 construct used the 5′ and 3′ sequences flanking *PTIWI09* (504 bp upstream and 236 bp downstream). FLAGHA-Ptiwi10 used the sequences flanking *DCL5* (352 upstream and 108 nt downstream). Frozen pellets from 1,200,000 *P. tetraurelia* cells, grown from a single cell transformed by microinjection with the designed constructs, were lysed in 2 mL of lysis buffer (composition given below) in a Dounce homogenizer. 1 mL of lysate was mixed with 50 μL of anti-HA affinity matrix (Roche) and incubated overnight with agitation (4 rpm) at 4°C. After this, beads were washed five times with immunoprecipitation (IP) buffer (composition given below). After the last wash, beads were resuspended in 1 mL of IP buffer. For western blotting, 100 μL of the resuspended beads and IP buffer was taken and centrifuged. Beads were boiled for 5 min in standard SDS loading dye. Mouse anti-HA (Y-11) (Santa Cruz Biotechnology) 1:500 was used. For extraction of associated RNAs, the supernatant was removed and mixed with 300 μL of Proteinase K buffer (1×, recipe below) and 1 μL of Proteinase K (10 mg/mL), followed by 20 min of incubation at 42°C. RNA was extracted with phenol:chloroform and then chloroform, and then precipitated and resuspended in 10 μL Mille-Q water. 1 μL of RNA was 5′ end-labeled and analyzed on a 16% polyacrylamide/7 M urea gel.

Lysis Buffer

Lysis buffer contained 50 mM Tris-Hcl (pH 8.0), 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 0.5% sodium deoxycholate, 1% Triton X-100, 1 × protease inhibitor complete tablet (Roche), 2 mM vanadyl ribonucleoside complex, and 10% glycerol.

IP Buffer

IP buffer contained 10% glycerol, 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 0.5% sodium deoxycholate, 1% Triton X-100, 1 × complete EDTA-free protease inhibitor tablet (Roche), and 2 mM vanadyl ribonucleoside complex (Sigma).

Proteinase K buffer (1×)contained 10 mM Tris (pH 7.5), 5 mM EDTA (pH 8.0), and 0.5% SDS.

Estimation of IES Retention and Examination of IRS Correlations

The following reference genomes were used in the IES analyses and read mapping: MAC, http://paramecium.cgm.cnrs-gif.fr/download/fasta/assemblies/ptetraurelia_mac_51.faa; MAC+IES, http://paramecium.cgm.cnrs-gif.fr/download/fasta/assemblies/ptetraurelia_mac_51_with_ies.fa (md5 checksums: dbcc54fb2987c8f60f8e765db7ed274c and 3e5b3fa65ebfaa484566a1ffddf20239, respectively).

IRSs were determined by ParTIES (Denby Wilkes et al., 2016). For each IES, ParTIES counts mapped reads with unexcised IESs (IES^+) and excised IESs (IES^-) to determine an IRS ($IRS = IES^+ \div (IES^+ + IES^-)$).

For the purpose of regression, for each IES, negative control IRS (from an empty vector control and non-developmental gene control; E.C.S., unpublished data) have been subtracted and zeroed when the subtraction yields a negative value; i.e., $IRS_{regression} = \max(IRS_{experiment} - \text{mean}(IRS_{controls}), 0)$. Three regressions are given for IRS pairs: an ordinary least-squares (OLS) regression, a LOWESS regression (non-parametric), and an orthogonal distance regression (ODR) with a linear function initialized by the slope and intercept from the OLS. OLS and ODR used standard SciPy functions. LOWESS regressions used the Python statsmodels library. Pearson's correlation coefficient was calculated for each pair of knockdown IRSs.

RNA Electrophoresis for Detection of Small RNA

Total RNA samples (5.6 μg) were 5′ end-labeled with γ-32P desoxyadenosintriphosphate (dATP) (0.4 MBq, Amersham) by the exchange reaction of T4 polynucleotide kinase (Thermo Scientific). To the RNA, 2 volumes of denaturing buffer (Formamide, Sigma: 6× loading dye [Thermo Scientific] = 9:1) was added. The radioactively labeled DNA was run on a 16% polyacrylamide/7 M urea gel.

Northern Blot

7 μg of RNA was separated by denaturing gel electrophoresis and transferred to Hybond N + membranes (GE Healthcare). The following probes were used: *PTIWI01*, 546 bp; *PTIWI09*, 546 bp; *PTIWI10*, 500 bp; *PTIWI11*, 551 bp, GFP, 828 bp entire sequence, 17S rRNA; Baudry et al., 2009). The 17S rRNA probe was labeled with γ-32P dATP (2 MBq); the other probes were labeled with α-32P dATP (2 MBq). Hybridization of all probes was performed in Church buffer overnight at 55°C. Blots were washed twice for 15 min with 2× saline sodium citrate (SSC), 0.1% SDS at 55°C prior to exposure. 17S rRNA probe hybridization and washing was performed at 55°C. Blots were washed twice for 15 min with 2× SSC, 0.1% SDS at 55°C prior to exposure. The primers used for the probes are listed in the Supplemental Experimental Procedures.

Dot Blot

To validate the DNA microinjection, 400 μL of cells at a concentration of 1,000 cells/mL were lysed and then bound to a Hybond N + membrane (GE Healthcare). To analyze transposon retention, 500 ng of DNA was used. *Sardine*, *Thon*, and Actin-specific probes were labeled with α-32P dATP (2 MBq), and hybridization and washing were performed at 60°C. The primers used for probes are listed in the Supplemental Experimental Procedures.

Sequence of FLAGHA Adjusted to *Paramecium tetraurelia* Codon Usage

The sequence of FLAGHA adjusted to *Paramecium tetraurelia* codon usage was ATGGATTATAAAGATCATGATGGAGACTACAAGGACCACGACATTGAT TATAAAGATGATGATGATAAATATCCATACGATGTTCCAGATTATGCTCAT.

Phylogenetic Tree

Multiple sequence alignment was performed by Multiple Alignment using Fast Fourier Transform (MAFFT) v7.017 (Katoh et al., 2002) (G-INS-i algorithm, BLOSUM62 scoring matrix, and a gap opening penalty of 1.53). 571 conserved alignment columns were automatically selected with TrimAl v.1.3 and the "-automated1" option. To create a phylogeny, PhyML Online was used with default settings and performed with 1,000 bootstrap (Guindon et al., 2010). Phylogeny visualization was done using Interactive Tree of Life (iTOL) (Letunic and Bork 2007). The following abbreviations were used: ARTH, *Arabidopsis thaliana*; CAEL, *Caenorhabditis elegans*; DIDI, *Dictyostelium discoideum*; DRME, *Drosophila melanogaster*; HOSA, *Homo sapiens*; OXTR, *Oxytricha trifallax*; PATE, *Paramecium tetraurelia*; SCPO, *Schizosaccharomyces pombe*; TETH, *Tetrahymena thermophila*; and TRBU, *Trypanosoma brucei*. The accession numbers of proteins used for the phylogenetic tree are listed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession numbers for the raw DNA sequencing and sRNA-seq data reported in this paper are European Nucleotide Archive: ERS1656548 (Ptiwi01/09-KD_DNA), ERS1656549 (Ptiwi10/11-KD_DNA), ERS1656550 (FLAGHA-Pt09_RIP_early), ERS1656551 (FLAGHA-Pt09-RIP_late), ERS1656552 (FLAGHA-Pt10-late+), ERS1658936 (Ptiwi01/09-KD_smallRNA_early), ERS1658937 (Ptiwi01/09-KD_smallRNA_late), ERS1658938 (Ptiwi01/09-KD_smallRNA_late+), ERS1658939 (Ptiwi10/11-KD_smallRNA_early), ERS1658940 (Ptiwi10/11-KD_smallRNA_late), ERS1658941 (Ptiwi10/11-KD_smallRNA_late+), ERS1658942 (EV-KD_smallRNA_early), ERS1658943 (EV-KD_smallRNA_late), and ERS1658944 (EV-KD_smallRNA_late+).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.06.050.

AUTHOR CONTRIBUTIONS

D.I.F. performed most of the experiments. D.I.F. and E.C.S. performed the bioinformatics analyses. M.F.K. performed cloning and imaging of GFP-Ptiwi01. P.Y.S. cloned FLAGHA-Ptiwi09. M.N. supervised the project. D.I.F., E.C.S., and M.N. wrote the manuscript.

ACKNOWLEDGMENTS

This research was supported by European Research Council Grants 260358 "EPIGENOME" and 681178 "G-EDIT", Swiss National Science Foundation Grants 31003A_146257 and 31003A_166407, and grants from the National Center of Competence in Research RNA and Disease. We thank Nasikhat Stahlberger for technical support and members of the Nowacki laboratory for discussions. We especially thank Simran Bhullar for discovering the effect of *PGM*-KD on *PTIWI* gene expression. We also thank the Pillai lab for technical help with RIP experiments.

Received: April 28, 2017
Revised: June 2, 2017
Accepted: June 20, 2017
Published: July 11, 2017

REFERENCES

Allen, S.E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., and Nowacki, M. (2017). Circular concatemers of ultra-short DNA segments produce regulatory RNAs. Cell *168*, 990–999.e7.

Arnaiz, O., Cain, S., Cohen, J., and Sperling, L. (2007). ParameciumDB: a community resource that integrates the Paramecium tetraurelia genome sequence with genetic data. Nucleic Acids Res. *35*, D439–D444.

Arnaiz, O., Goût, J.F., Bétermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E., and Sperling, L. (2010). Gene expression in a paleopolyploid: a transcriptome resource for the ciliate Paramecium tetraurelia. BMC Genomics *11*, 547.

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. PLoS Genet. *8*, e1002984.

Baranasic, D., Oppermann, T., Cheaib, M., Cullum, J., Schmidt, H., and Simon, M. (2014). Genomic characterization of variable surface antigens reveals a telomere position effect as a prerequisite for RNA interference-mediated silencing in Paramecium tetraurelia. MBio *5*, e01328.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase

involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev. *23*, 2478–2483.

Beisson, J., Bétermier, M., Bré, M.H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr., and Sperling, L. (2010). Paramecium tetraurelia: the renaissance of an early unicellular model. Cold Spring Harb. Protoc. *2010*, emo140.

Bouhouche, K., Gout, J.F., Kapusta, A., Bétermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. Nucleic Acids Res. *39*, 4249–4264.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089–1103.

Cora, E., Pandey, R.R., Xiol, J., Taylor, J., Sachidanandam, R., McCarthy, A.A., and Pillai, R.S. (2014). The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA *20*, 773–781.

Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat. Rev. Genet. *12*, 19–31. http://www.ncbi.nlm.nih.gov/pubmed/21116305.

Denby Wilkes, C., Arnaiz, O., and Sperling, L. (2016). ParTIES: a toolbox for Paramecium interspersed DNA elimination studies. Bioinformatics *32*, 599–601.

Duharcourt, S., Keller, A.M., and Meyer, E. (1998). Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in Paramecium tetraurelia. Mol. Cell. Biol. *18*, 7075–7085.

Fang, W., Wang, X., Bracht, J.R., Nowacki, M., and Landweber, L.F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell *151*, 1243–1255.

Galvani, A., and Sperling, L. (2002). RNA interference by feeding in Paramecium. Trends Genet. *18*, 11–12.

Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23–35.

Götz, U., Marker, S., Cheaib, M., Andresen, K., Shrestha, S., Durai, D.A., Nordström, K.J., Schulz, M.H., and Simon, M. (2016). Two sets of RNAi components are required for heterochromatin formation in trans triggered by truncated transgenes. Nucleic Acids Res. *44*, 5908–5923.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Haase, A.D. (2016). A Small RNA-Based Immune System Defends Germ Cells against Mobile Genetic Elements. Stem Cells Int. *2016*, 7595791.

Katoh, M., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Lepère, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. Nucleic Acids Res. *37*, 903–915.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics *23*, 127–128.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.

Marker, S., Le Mouël, A., Meyer, E., and Simon, M. (2010). Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in Paramecium tetraurelia. Nucleic Acids Res. *38*, 4092–4107.

Miyoshi, T., Ito, K., Murakami, R., and Uchiumi, T. (2016). Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. Nat. Commun. *7*, 11846.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368–374.

Nowacki, M., Zagorski-Ostoja, W., and Meyer, E. (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear cross-talk in Paramecium tetraurelia. Curr. Biol. *15*, 1616–1628.

Ruiz, F., Vayssié, L., Klotz, C., Sperling, L., and Madeddu, L. (1998). Homology-dependent gene silencing in Paramecium. Mol. Biol. Cell *9*, 931–943.

Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. Dev. Cell *28*, 174–188.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468–474.

Cell Reports, Volume *20*

Supplemental Information

Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA

Dominique I. Furrer, Estienne C. Swart, Matthias F. Kraft, Pamela Y. Sandoval, and Mariusz Nowacki

Supplemental Data

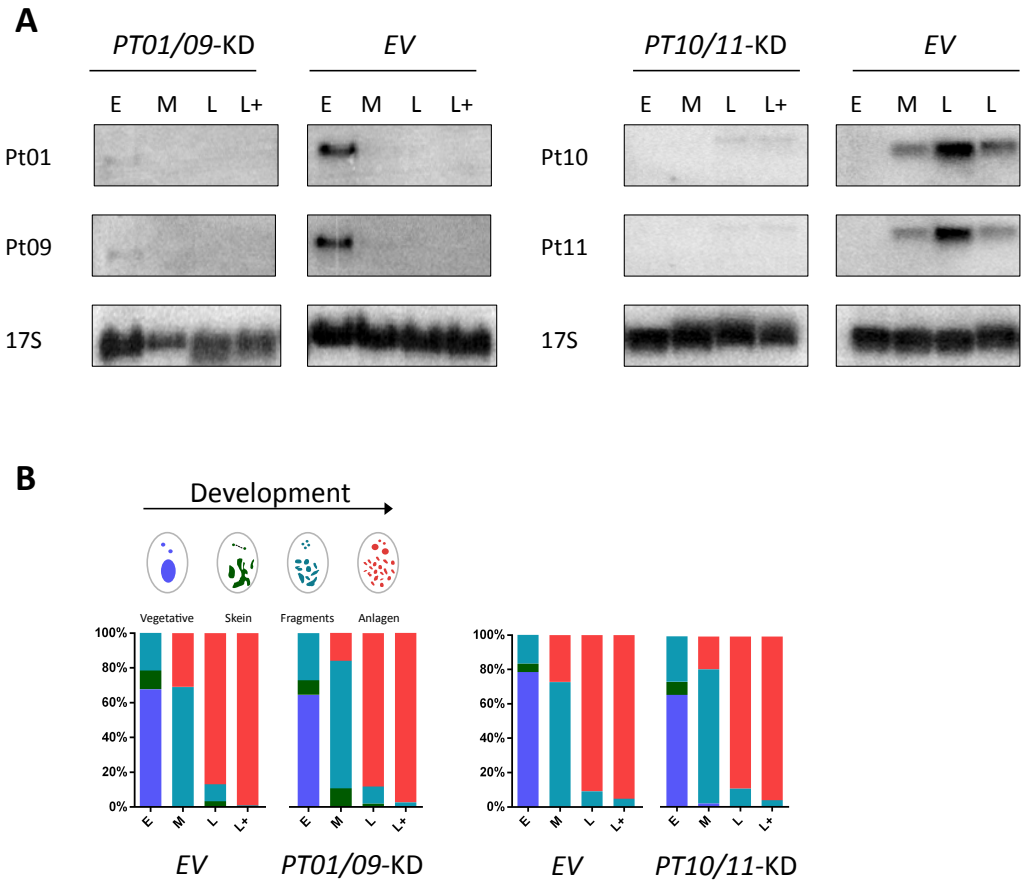


Figure S1. Effects of *PTIWI01/09*-KD and *PTIWI10/11*-KD on IES and transposon excision, Related to Figure 3 and 4.

A) Northern blot analysis of *PTIWI* expression in total RNA obtained from developing cells after *PTIWI01/09*-KD and *PTIWI10/11*-KD. Probes against each of the silenced genes were used; a probe complementary to 17S rRNA is the loading control. Note that due to the fact that the ohnologs of PTIWI01 and PTIWI09 (PTIWI10 and PTIWI11, respectively) are highly similar to each other, we did not design probes that do not also bind to these ohnologs. **B)** Distribution of cytological stages of samples collected during cell development from empty vector (EV), *PTIWI01/09*-KD and *PTIWI10/11*-KD cells. A schematic drawing of the nuclear morphology used to characterization the stages is shown above. Bars indicate percentages of cells in a specific stage for each of the collected samples. See Experimental Procedures for additional details. Blue: Intact old MAC. Green: Cells with the old MAC displaying the characteristic skein formation. Turquoise: Old MAC is completely fragmented. Red: Visible new MACs.

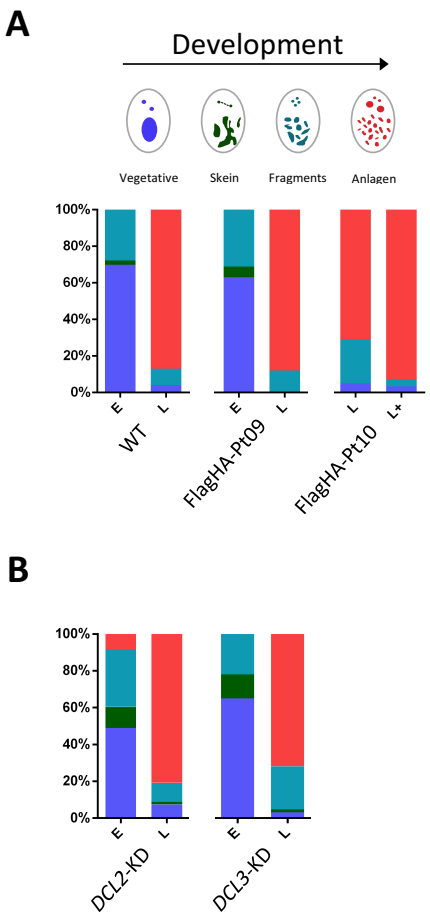


Figure S2. Small RNAs from different time points obtained from RNA immunoprecipitation (RIP) of FlagHA-Ptiwi09 and FlagHA-Ptiwi10 proteins expressed in *Paramecium* cells, Related to Figure 5.
A) Distribution of cytological stages of samples collected during cell development from (wild type, WT), FlagHA-Ptiwi09 and FlagHA-Ptiwi10 cells used during RIP. A schematic drawing of the nuclear morphology used to characterization the stages is shown above. Bars indicate percentages of cells in a specific stage for each of the collected samples. See Experimental Procedures for additional details. Blue: Intact old MAC. Green: Cells with the old MAC displaying the characteristic skein formation. Turquoise: Old MAC is completely fragmented. Red: Visible new MACs. **B)** Distribution of cytological stages of samples collected during cell development from FlagHA-Ptiwi09 cells in *DCL2*-KD and *DCL3*-KD backgrounds.

Table S1. IES retention scores, **Related to Figure 1 and 3,**

IES		**Retention score**						
Classical IES name	IES name in reference MAC+ IES assembly	*Maternal control score*	*EV*	*PGM*-KD	*DCL2/3*-KD	*DCL5*-KD	*PT01/09*-KD	*PT10/11*-KD
51G11	IESPGM.PTET51.1.51.451201		0.03	**0.80**	0.05	**0.20**	0.02	**0.20**
51G1413	IESPGM.PTET51.1.51.452624	0.00	0.01	**0.78**	0.01	0.02	0.00	0.00
51G1832	IESPGM.PTET51.1.51.453043	0.00	0.00	**0.73**	0.00	0.00	0.01	0.01
51G2832	IESPGM.PTET51.1.51.454043	**0.87**	0.00	**0.77**	**0.23**	0.00	**0.41**	0.00
51G4404	IESPGM.PTET51.1.51.455615	**0.97**	0.00	**0.84**	**0.60**	0.01	**0.77**	0.01
51G6447	IESPGM.PTET51.1.51.457658		0.00	**0.78**	0.04	0.00	0.01	0.01
51A712	IESPGM.PTET51.1.106.281631	**0.34**	0.00	**0.78**	0.07	0.06	**0.18**	0.08
51A1835	IESPGM.PTET51.1.106.284157	0.00	0.00	**0.80**	0.00	0.00	0.00	0.00
51A2591	IESPGM.PTET51.1.106.284913	**0.99**	0.00	**0.91**	**0.54**	0.00	**0.66**	0.00
51A4404	IESPGM.PTET51.1.106.286750	0.00	0.00	**0.83**	0.00	0.00	0.00	0.00
51A4578	IESPGM.PTET51.1.106.286924	0.00	0.00	**0.78**	0.05	0.00	0.03	0.00
51A6435	IESPGM.PTET51.1.106.288781	0.00	0.00	**0.77**	0.00	0.00	0.00	0.00
51A6649	IESPGM.PTET51.1.106.288995	**0.58**	0.00	**0.81**	**0.56**	0.01	**0.48**	0.00
Dcl5d-01	IESPGM.PTET51.1.8.257314		0.02	**0.75**	0.03	**0.50**	0.07	**0.25**
Dcl5d-02	IESPGM.PTET51.1.16.506105		0.00	**0.95**	0.02	**0.38**	0.05	**0.51**
Dcl5d-03	IESPGM.PTET51.1.39.60104		0.00	**0.74**	0.00	**0.37**	0.00	**0.25**
Dcl5d-04	IESPGM.PTET51.1.168.68297		0.00	**0.69**	0.05	**0.37**	0.06	**0.22**
A-10	IESPGM.PTET51.1.106.282313	0.00	0.00	**0.87**	0.00	0.00	0.00	0.00
A1416	IESPGM.PTET51.1.106.283738	0.00	0.00	**0.76**	0.00	0.00	0.00	0.00

.

Supplemental Experimental Procedures

Primers used in this study

Target		Primers	
IESs		Forward Primer	Reverse Primer
Classical IES name	IES name in reference MAC+ IES genome assembly		
51G11	IESPGM.PTET51.1.51.451201	ATCATAAGATTGATATCTTCTCCCTTCTCC	ACTTGCTACTAAAGCAAGAAACATTGAGAG
51G1413	IESPGM.PTET51.1.51.452624	GAAGCTGCTTGTGTTAAGAATTCTACTGG	GCATCCAGCACTAGTTGAATTTACTGTAC 0.01
51G1832	IESPGM.PTET51.1.51.453043	CTATAACTCTTGAAGCTGCTTGTAATATG	TTGTCAATGAGCCATTAACAGTTGCTGGAT
51G2832	IESPGM.PTET51.1.51.454043	GCTATAACTCTTGAAGCTGCTTGTAATATG	TTGTCAATGAGCCATTAACAGTTGCTGGAT
51G4404	IESPGM.PTET51.1.51.455615	CTGTTGCTACACATTGTGCATATGTTACT	GCTGTAAGATTAACATTGAGCATGATCAAG
51G6447	IESPGM.PTET51.1.51.457658	AATGCATCAAATGTAGTAACTACTCCTGCT	AATTTGTAAAGTATCCAGCGCAGGCAG
51A712	IESPGM.PTET51.1.106.281631	TTTGTCAAAAAGACATGTATCAAAATGCAG	TAGAATACTAAGAGATTCAATACAACAAAC
51A1835	IESPGM.PTET51.1.106.284157	TAATGTATTGATAAGGCTTGCTCTACAGCC	ATCTAACATCCTTGAATAGTTACTGATCC
51A2591	IESPGM.PTET51.1.106.284913	ATGTGTTTGGACTGGATTGGCATGTAGAAG	GATGTAGCATAACATTTATCAACAATCCAT
51A4404	IESPGM.PTET51.1.106.286750	TGGAATAGTGCTGCATCACCAGCTGCTTGC	CCAGTTATTGAACTGCAACTTACTGCAGTG
51A4578	IESPGM.PTET51.1.106.286924	CCTGCAGTAAGTTGCAGTTCAATAACTGG	TGTAGTCTTAAAATCTTAGCATGTTGTACC
51A6435	IESPGM.PTET51.1.106.288781	CAAATTGTGTCACTAGAGGTACATGTTTCC	GCGACATCAATAGTAACAGCTGAGCATGAG
51A6649	IESPGM.PTET51.1.106.288995	ACTGCACCTCTAACTTTAACAAGCGAAGCA	CAGCAGTACATCCAGCTCTCTAAGTTTAGC
Dcl5d-01	IESPGM.PTET51.1.8.257314	CCAGTCTTTATAACTCCAAATATACTAATGTT AATTGCC	CTTGCTGGTTGAATATCAATTGAAAAATCTTGATG
Dcl5d-02	IESPGM.PTET51.1.16.506105	TTTTCATACTCATCCTCACCCTACTCCC	ATAATATAAACTTTGAAGCCCCTGAAGCCG
Dcl5d-03	IESPGM.PTET51.1.39.60104	TTCATCGTTCTCAAAATGGATGCCC	ATCAATATGAATTTTATTTGTTTGTTTATAAGCGTCTGG
Dcl5d-04	IESPGM.PTET51.1.168.68297	TTAAGGTACTTCTTCATCCATAGTCAGCC	AACAATATGAGTTAGAAATTTAAAATCTTGGAGGAATCC
Silencing constructs			
PTIWI01/ PTIWI09		ACTTCCATGAGATGAGATTGAACCC	TACTTTGGCTTTATTGGCTTCAGC
PTIWI10		TCTATAAAGTGAATAAAGTGAGTAGGGAG	TCTTTGTGAAATCAGATAATAATCATCC
PTIWI11		TCTAAACTATGAAATAACTATCCACAGTC	TGTTTCAATGACAGTTTTAGCTTAAGAG
Probes			
PTIWI01/ PTIWI09		ATATTATTCTCAAGTTGTTACATCTGG	ATATCTAACAGGAGCAGGG
PTIWI10		TACAGATATCAAGTAAGATATGCATGTC	AAGAGTCGATGTTGATGTAATACTTC
PTIWI11		ATCTACAGATATCAAGTAACCCATCCC	AATTCCTCAACTCCTCGTACATAGTG
Sardine		GAACGCACTCGAAATGCAAGTGCTGC	GCCAGCTCATATAGGATAAAGAGTCTG
Thon		AGCGAACAGTTTATGGTAGTACCATTC	GACACTGAGTAGCTTGCCAATTTTCAA
Actin		ATGTCTGAAGAACACCCAGCAGTCGTTATTG	ATCCCAATTATTGACAATACCATTATCAATTG
17s rRNA		ACC CGT GAC TGC CAT GGT AGT CCA ATA CA	-

Accession number of proteins used for the Phylogenetic tree

Arabidopsis thaliana:					
AGO1:	NP_175274.1	AGO2:	NP_174413.2	AGO3:	NP_174414.1
AGO4:	NP_565633.1	AGO5:	NP_850110.1	AGO6:	NP_180853.2
AGO7:	NP_177103.1	AGO8:	NP_197602.2	AGO9:	NP_197613.2
AGO10:	NP_199194.1				
Caenorhabditis elegans:					
WAGO-1:	NP_492045.1	PPW-2:	NP_491535.1	WAGO-4:	NP_496751.1
WAGO-5:	NP_495151.3	SAGO-2	NP_490758.1	PPW-1:	NP_740835.1
SAGO-1	NP_504610.1	HRDE-1:	NP_497834.1	WAGO-10:	NP_503177.1
WAGO-11:	NP_494593.1	NRDE-3	NP_508092.1	ERGO-1:	NP_503362.2
PRG-1:	NP_492121.1	PRG-2:	NP_500994.1		
Dictyostelium discoideum					
AgnA:	XP_643218.1	AgnB:	XP_635708.1	AgnC:	XP_645445.1
AgnD:	XP_001134555.1	AgnE:	XP_636288.1		
Drosophila melanogaster					
Ago1:	NP_725341.1	Ago2:	NP_648775.1	Ago3:	NP_001036627.2
Aub:	NP_476734.1	Piwi:	NP_476875.1		
Homo sapiens					
AGO1:	NP_001304051.1	AGO2:	NP_036286.2	AGO3:	NP_079128.2
AGO4:	NP_060099.2	PIWIL1:	NP_004755.2	PIWIL2:	NP_060538.2
PIWIL3:	NP_001008496.2	PIWIL4:	NP_689644.2		
Oxytricha trifallax					
Otiwi01:	AEX87959.1	Otiwi02:	AEX87964.1	Otiwi03:	AEX87965.1
Otiwi04:	AEX87966.1	Otiwi05:	AEX87967.1	Otiwi06:	AEX87968.1
Otiwi07:	AEX87969.1	Otiwi08:	AEX87970.1	Otiwi09:	AEX87971.1
Otiwi10:	AEX87960.1	Otiwi11:	AEX87961.1	Otiwi12:	AEX87962.1
Otiwi13:	AEX87963.1				
Paramecium tetraurelia					
Ptiwi01:	XP_001456124.1	Ptiwi02:	CAI44470.1	Ptiwi03:	XP_001442589.1
Ptiwi05:	XP_001437215.1	Ptiwi06:	XP_001452498.1	Ptiwi07:	CAI39074.1
Ptiwi08:	XP_001455053.1	Ptiwi09:	XP_001454540.1	Ptiwi10:	XP_001440517.1
Ptiwi11:	XP_001453643.1	Ptiwi12:	XP_001446065.1	Ptiwi13:	CAI39067.1
Ptiwi14:	CAI39066.1	Ptiwi15:	XP_001428017.1		
Schizosaccharomyces pombe					
Ago1:	NP_587782.1				
Tetrahymena thermophila					
Twi01:	BAC02573.1	Twi02:	XP_001015192.1	Twi07:	XP_001032516.1
Twi08:	ABI15747.1	Twi09:	XP_001007088.2	Twi10:	XP_001027583.2
Twi11:	XP_001011123.3	Twi12:	XP_001013933.2		
Trypanosoma brucei					
AGO1:	AAR10810.1	TBIWI:	AAR10811.1		